

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 21:28:32 : Search time 536 Seconds

(Without alignments)
8530.060 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043

Sequence: 1 gaattccattgtgttgta.....tgggatacttaaaaaaaaaa 2043

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1504479 seqs, 1118970152 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2975614

Minimum DB seq length: 15

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCF_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1547	75.7	2032	9 US-09-816-825-1	Sequence 1, Appl1
2	1547	75.7	2032	13 US-10-007-262-2	Sequence 2, Appl1
3	403	19.7	505	10 US-09-998-598-2595	Sequence 2595, Ap
4	403	19.7	517	10 US-09-919-580-194	Sequence 194, Ap
5	114	5.6	389	10 US-09-919-580-264	Sequence 264, App
6	59	2.9	2544	9 US-09-927-602-1	Sequence 1, Appl1
7	59	2.9	48436	9 US-09-927-602-1	Sequence 38, Appl1
8	37	1.8	37	9 US-09-816-825-7	Sequence 7, Appl1
9	37	1.8	37	13 US-10-007-262-7	Sequence 7, Appl1
10	22	1.1	167	10 US-09-867-701-9458	Sequence 9458, Ap
11	22	1.1	171	10 US-09-867-701-9067	Sequence 9067, Ap
12	22	1.1	187	10 US-09-867-701-9185	Sequence 9185, Ap
13	22	1.1	191	10 US-09-867-701-9284	Sequence 9284, Ap
14	20	1.0	398	10 US-09-860-352-11099	Sequence 11099, A
15	20	1.0	370	13 US-10-027-632-140660	Sequence 140660, Sequence 334, App
16	20	1.0	11102	14 US-10-205-823-334	Sequence 334, App

17	20	1.0	180557	13 US-10-003-806-6	Sequence 6, Appl1
18	20	1.0	180557	13 US-10-003-806-9	Sequence 9, Appl1
19	19	0.9	341	10 US-09-867-701-8841	Sequence 8841, Ap
20	19	0.9	341	10 US-09-867-701-8888	Sequence 8888, Ap
21	19	0.9	349	9 US-09-759-143-649	Sequence 649, App
22	19	0.9	349	9 US-09-780-669-649	Sequence 649, App
23	19	0.9	349	9 US-09-822-827-649	Sequence 649, App
24	19	0.9	349	10 US-09-895-793-649	Sequence 649, App
25	19	0.9	349	10 US-09-895-814-649	Sequence 649, App
26	19	0.9	349	13 US-10-012-896-649	Sequence 649, App
27	19	0.9	405	10 US-09-867-701-9634	Sequence 9634, Ap
28	19	0.9	544	13 US-10-027-632-118925	Sequence 118925, Sequence 11785,
29	19	0.9	786	13 US-10-027-632-118925	Sequence 26, Appl1
30	19	0.9	2874	13 US-10-027-632-111785	Sequence 52, Appl1
31	19	0.9	3273	12 US-09-871-392-26	Sequence 944, Ap
32	19	0.9	3274	13 US-10-002-600-52	Sequence 322, App
33	19	0.9	4224	9 US-09-815-242-944	Sequence 2201, Ap
34	19	0.9	4537	13 US-10-108-605-322	Sequence 147, App
35	19	0.9	6546	12 US-10-017-161-2201	Sequence 6217, Ap
36	19	0.9	8742	9 US-09-764-891-6217	Sequence 7, Appl1
37	19	0.9	10925	14 US-10-032-393-7	Sequence 2257, Ap
38	19	0.9	62944	10 US-09-954-456-2257	Sequence 243, App
39	19	0.9	75899	9 US-09-854-883-243	Sequence 1, Appl1
40	19	0.9	684973	10 US-09-263-959-1	Sequence 17017, A
41	19	0.9	25	14 US-10-098-2638-17017	Sequence 8, Appl1
42	18	0.9	47	9 US-09-816-825-8	Sequence 765, App
43	18	0.9	47	13 US-10-007-262-8	
44	18	0.9	129	9 US-09-759-143-765	
45	18	0.9			

ALIGNMENTS

RESULT 1	US-09-816-825-1	Application US/09816825
Sequence 1, Application US/09816825	Patent No. US20010051370A1	
GENERAL INFORMATION:		
APPLICANT: Bistrup, Annette		
APPLICANT: Rosen, Steven D.		
APPLICANT: Hemmerich, Stefan		
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3		
FILE REFERENCE: 6510-107CON		
CURRENT APPLICATION NUMBER: US/09/816, 825		
CURRENT FILING DATE: 2001-03-22		
PRIOR APPLICATION NUMBER: 09/045, 284		
PRIOR FILING DATE: 1998-03-20		
NUMBER OF SEQ ID NOS: 9		
SOFTWARE: FastSeq for Windows Version 4.0		
SEQ ID NO 1		
LENGTH: 2032		
TYPE: DNA		
ORGANISM: Homo sapiens		
US-09-816-825-1		
Query Match	75.7%: Score 1547; DB 9; Length 2032;	
Best Local Similarity	99.9%: Pred. No. 0;	
Matches 1597; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
QY	106 CAAGGCTTCCTCAGCACAATGCTACCTCTTAAATGAAGCTCTGCTTTCT	165
Db	160 CAAGGCTTCCTCAGCACAATGCTACCTCTTAAATGAAGCTCTGCTTTCT	219
QY	166 GGTTCCTCAGTGGCCATCTTGGCTTATCTTCACATGACGACCAATCAGCTC	225
Db	220 GGTTCCTCAGTGGCCATCTTGGCTTATCTTCACATGACGACCAATCAGCTC	279
QY	226 CCGTCTATGAGGACAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	285
Db	280 CCGTCTATGAGGACAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	339
QY	286 TGGCTCTCTTCTTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	345

```

Db      ||| 1420 TCTGTAGACCTTAACTACATGTCGTGGGTATACACTGAGTGTGATGTGCCACAG 1479
QY      ||| 1426 TGCTCAAGCAGAGACTTTTGTCCATGCTGTGTCTGAAACAGACTGGGAACT 1485
Db      ||| 1480 TGCTCAAGCAGAGACTTTTGTCCATGCTGTGTCTGAAACAGACTGGGAACT 1539
QY      ||| 1486 TATGTAGACGACATATCCACAGTGAAGAGGATATGCTCTTCTTCTTCTTCT 1545
Db      ||| 1540 TATGTAGACGACATATCCACAGTGAAGAGGATATGCTCTTCTTCTTCTTCT 1599
QY      ||| 1546 TCTGTCTGGGACACTTTCAGACATTTTGTGGCTGGAGGCTTATTAAGCAGACAGT 1605
Db      ||| 1600 TCTGTCTGGGACACTTTCAGACATTTTGTGGCTGGAGGCTTATTAAGCAGACAGT 1659
QY      ||| 1606 ATCAGTGAATTTATTCATTAACCTCCCTGTCACATCTTCCCAATGGGAAATGATCT 1665
Db      ||| 1660 ATCAGTGAATTTATTCATTAACCTCCCTGTCACATCTTCCCAATGGGAAATGATCT 1719
QY      ||| 1666 TTCACCAAGAGCTCACACAGATTTTCCACAGATGC 1703
Db      ||| 1720 TTCACCAAGAGCTCACACAGATTTTCCACAGATGC 1757

RESULT 2
US-10-007-262-2
; Sequence 2, Application US/10007262
; Publication No. US20020164748A1
; GENERAL INFORMATION:
; APPLICANT: Blistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerlich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/10/007,262
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens
; US-10-007-262-2

Query Match          75.7%; Score 1547; DB 13; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Db      ||| 106 CAAGGTCTCCACTTACAGCACAATGCTACTGCTTAAAAAATGAAGCTCTGCTTCT 165
QY      ||| 160 CAAGGTCTCCACTTACAGCACAATGCTACTGCTTAAAAAATGAAGCTCTGCTTCT 219
Db      ||| 166 GGTTCCTCAAGATGGCCATCTTGGCTTATTTCCACATGTATACGCCACAATACAGCTC 225
QY      ||| 220 GGTTCCTCAAGATGGCCATCTTGGCTTATTTCCACATGTATACGCCACAATACAGCTC 279
Db      ||| 226 CCTGTATGAAGGACACGCCAGGAGATGACAGTGTGTCTCTTCTTCCGAGGCTC 285
QY      ||| 280 CCTGTATGAAGGACACGCCAGGAGATGACAGTGTGTCTCTTCTTCCGAGGCTC 339
Db      ||| 286 TGCTCTCTCTTCTTGTGGGCAAGCTTTTGGGCAAGCAGATGTTTCTTACGTATGA 345
QY      ||| 340 TGCTCTCTCTTCTTGTGGGCAAGCTTTTGGGCAAGCAGATGTTTCTTACGTATGA 399
Db      ||| 346 GCCCGCTGGACAGTGTGTGATGATCTTCAAGCAGACACCGCTGTGATGTGACATGGC 405
QY      ||| 400 GCCCGCTGGACAGTGTGTGATGATCTTCAAGCAGACACCGCTGTGATGTGACATGGC 459
Db      ||| 406 TGTGCGGATCTGATACGAGGCGCTTCTTGTGCGACATGAGAGCTTTGATGAGCTTACAT 465
QY      ||| 460 TGTGCGGATCTGATACGAGGCGCTTCTTGTGCGACATGAGAGCTTTGATGAGCTTACAT 519

```

QY 466 GGAACCTGTCCTCCCGGAGACAGTCACGCTCTTTCAGTGGGAGAACAGCCGGGCTGTG 525
DB 520 GGAACCTGTCCTCCCGGAGACAGTCACGCTCTTTCAGTGGGAGAACAGCCGGGCTGTG 579
QY 526 TTTCGACCTGCTGCTGTGATCATCCCAAGATGAATCATCCCGGGCTCATCTGAG 585
DB 580 TTTCGACCTGCTGCTGTGATCATCCCAAGATGAATCATCCCGGGCTCATCTGAG 639
QY 586 GCTCTGTGAGTACAGCCCTTGAAGTGTGGAGAGGCTGGCGCTCTACAGCA 645
DB 640 GCTCTGTGAGTACAGCCCTTGAAGTGTGGAGAGGCTGGCGCTCTACAGCA 699
QY 646 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACGCTCCCTACCCGCTGTGAAGA 705
DB 700 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACGCTCCCTACCCGCTGTGAAGA 759
QY 706 CCCCTCCCTCAACCTGATATCGTGACCTGTGCTGGGAGCCCGGGGCTGTTCGTC 765
DB 760 CCCCTCCCTCAACCTGATATCGTGACCTGTGCTGGGAGCCCGGGGCTGTTCGTC 819
QY 766 CCGAGAACGCAAGAGAGATCATGATTGACAGTGGATGTGGGGGAGCATGA 825
DB 820 CCGAGAACGCAAGAGAGATCATGATTGACAGTGGATGTGGGGGAGCATGA 879
QY 826 GCAAAAACCTCAAGAGAGAGACCAACCTTACTATGTATGAGGTCTCATCTGCAAGCA 885
DB 880 GCAAAAACCTCAAGAGAGAGACCAACCTTACTATGTATGAGGTCTCATCTGCAAGCA 939
QY 886 GCTGGAATCTCAAGACCATCCAGTCTTGGCCCAAGGCTCGAGAGACCTACCTGCT 945
DB 940 GCTGGAATCTCAAGACCATCCAGTCTTGGCCCAAGGCTCGAGAGACCTACCTGCT 999
QY 946 TGTGCGTATGAGAGACCTGGCTGAGGCCCTGTGGCCAGACTCCGAATGTGATTT 1005
DB 1000 TGTGCGTATGAGAGACCTGGCTGAGGCCCTGTGGCCAGACTCCGAATGTGATTT 1059
QY 1006 CGTGGATGTGAATTTTGGCCCATCTTTCAGACCTGGGTGATTAATCATCCAGGCA 1065
DB 1060 CGTGGATGTGAATTTTGGCCCATCTTTCAGACCTGGGTGATTAATCATCCAGGCA 1119
QY 1066 GGGGATGGGTGACACCGCTTTCACACCAATGCCAGGATGCCCTTAATGTCTCCAGGC 1125
DB 1120 GGGGATGGGTGACACCGCTTTCACACCAATGCCAGGATGCCCTTAATGTCTCCAGGC 1179
QY 1126 TTGGCGTGTGCTTGGCCATGATAAAGTTTCTGACTTCAAGAACCTGTGGCATGC 1185
DB 1180 TTGGCGTGTGCTTGGCCATGATAAAGTTTCTGACTTCAAGAACCTGTGGCATGC 1239
QY 1186 CATGAATTTGCTGGGCTACCGCCACGTCAATCTGAACAGAAAGAAACCTGTGCT 1245
DB 1240 CATGAATTTGCTGGGCTACCGCCACGTCAATCTGAACAGAAAGAAACCTGTGCT 1299
QY 1246 GGAATCTTCTGTCTACTGGAAGTCTCCCTGAGCAATCCACTAAGAGGTTGAGAAGCTT 1305
DB 1300 GGAATCTTCTGTCTACTGGAAGTCTCCCTGAGCAATCCACTAAGAGGTTGAGAAGCTT 1359
QY 1306 TGTGCGACCTGGGTGAGCTGCTGATGATTTCTCTGAATGCTTCTAGGCTTGCCTACA 1365
DB 1360 TGTGCGACCTGGGTGAGCTGCTGATGATTTCTCTGAATGCTTCTAGGCTTGCCTACA 1419
QY 1366 TCTCTGAGCTTAACTACATGTCTGGGTATCAACATGATGATGATGATGATGATGATG 1425
DB 1420 TCTCTGAGCTTAACTACATGTCTGGGTATCAACATGATGATGATGATGATGATGATG 1479
QY 1426 TGTCTAAGCAAGAGACTTTTGTGTCTATGCTGTGTCTAAGAAACAGACTGGGAACT 1485
DB 1480 TGTCTAAGCAAGAGACTTTTGTGTCTATGCTGTGTCTAAGAAACAGACTGGGAACT 1539
QY 1486 TATGTGAGCAAGCAATCCCAAGAGTGAAGAGGATTTGCTTCTTCTTCTGATCT 1545
DB 1540 TATGTGAGCAAGCAATCCCAAGAGTGAAGAGGATTTGCTTCTTCTTCTGATCT 1599

QY 1546 TCTGTCTGGGACAGCTTACAGACTTGTGGCTTGAGGAGCTTAATTAAGCAGACAGT 1605
DB 1600 TCTGTCTGGGACAGCTTACAGACTTGTGGCTTGAGGAGCTTAATTAAGCAGACAGT 1659
QY 1606 ATCACTGGAATTTGATCCATAAAGCTCCCTGTCCATCATCTTCCCAATGGGAATGATCT 1665
DB 1660 ATCACTGGAATTTGATCCATAAAGCTCCCTGTCCATCATCTTCCCAATGGGAATGATCT 1719
QY 1666 TTCAACAAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
DB 1720 TTCAACAAAGAGCTCACAGCATTTTCCACAGAGATGC 1757

RESULT 3

US-09-598-2595
; Sequence 2595, Application US/0998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jlangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Mesgher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.561
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Cortix Invention Disclosure Database
; SEQ ID NO 2595
; LENGTH: 505
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-2595

Query Match 19.7%; Score 403; DB 10; Length 505;
Best Local Similarity 100.0%; Pred. No. 8.5e-203;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GCGTTTCTGCGACCTGGTGTGACGCTTCTGTGATGCTTGTGAGCCCTTGC 1360
DB 1 GCGTTTCTGCGACCTGGTGTGACGCTTCTGTGATGCTTGTGAGCCCTTGC 60
QY 1361 CTACATCTCTAGGCTTACTACATGCTGTGGGTATCACAGTGTGATGCTGTC 1420
DB 61 CTACATCTCTAGGCTTACTACATGCTGTGGGTATCACAGTGTGATGCTGTC 120
QY 1421 ACAGTGTCTCAAGAGAGACTTTGTGTCATGCTTGTCTAGAAAACAGACTGGG 1480
DB 121 ACAGTGTCTCAAGAGAGACTTTGTGTCATGCTTGTCTAGAAAACAGACTGGG 180
QY 1481 AACCTTATGTGAGAGACATCCACAGTGAAGAGGATTTCTTCTTCTTCTT 1540
DB 181 AACCTTATGTGAGAGACATCCACAGTGAAGAGGATTTCTTCTTCTTCTT 240
QY 1541 GATCTTCTGTCTGGGACAGACTTGTGAGACTTGTGGGCTGGAGGCTTAATACAGAC 1600
DB 241 GATCTTCTGTCTGGGACAGACTTGTGAGACTTGTGGGCTGGAGGCTTAATACAGAC 300
QY 1601 ACAGTATCAGTGAATGATCAATAACTCCCTGTCCATCTTGGCCAAATGGGATG 1660
DB 301 ACAGTATCAGTGAATGATCAATAACTCCCTGTCCATCTTGGCCAAATGGGATG 360
QY 1661 GATCTTTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
DB 361 GATCTTTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 403

RESULT 4

US-09-919-580-194
; Sequence 194, Application US/09919580
; Patent No. US20020110832A1
; GENERAL INFORMATION:

APPLICANT: Pyle, Ruth
APPLICANT: Xu, Jiangchun
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.552
CURRENT APPLICATION NUMBER: US/09/919,580
CURRENT FILING DATE: 2001-07-30
NUMBER OF SEQ ID NOS: 934
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 194
LENGTH: 517
TYPE: DNA
ORGANISM: Homo sapiens
US-09-919-580-194

Query Match 19.7%; Score 403; DB 10; Length 517;
Best Local Similarity 100.0%; Pred. No. 8.5e-203; Indels 0; Gaps 0;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GGCTTGGCCACCTGCTGTCAGCCTCAGTCACCTTCTCTGAATGCTTCTGAGCCTTGC 1360
DB 13 GGCTTGGCCACCTGCTGTCAGCCTCAGTCACCTTCTCTGAATGCTTCTGAGCCTTGC 72
QY 1361 CTACATCTCTGAGCCTTAATCAATGCTGCTGGGTATCAACAGTGTGAGTGTCTC 1420
DB 73 CTACATCTCTGAGCCTTAATCAATGCTGCTGGGTATCAACAGTGTGAGTGTCTC 132
QY 1421 ACACGTGCTCAAGCAGAGAGACTTGTGTCTTCATGCTTGTCTAGAAAACAGACGTGGG 1480
DB 133 ACACGTGCTCAAGCAGAGAGACTTGTGTCTTCATGCTTGTGTCTAGAAAACAGACGTGGG 192
QY 1481 AACCTTATGTGAGACACACATCCACAGTGAAGAGGATGCTTCTTCTTCTTCTT 1540
DB 193 AACCTTATGTGAGACACACATCCACAGTGAAGAGGATGCTTCTTCTTCTTCTTCT 252
QY 1541 GATCTTCTGCTGGGACAGACTTCAGAGACTTGTGGCCGGAGGCTATTAGACAGAC 1600
DB 253 GATCTTCTGCTGGGACAGACTTCAGAGACTTGTGGCCGGAGGCTATTAGACAGAC 312
QY 1601 ACAGTATCAGTGAATGTATCCATAAACCTCCCTGTCACATCTTCCCAATGGGAGATG 1660
DB 313 ACAGTATCAGTGAATGTATCCATAAACCTCCCTGTCACATCTTCCCAATGGGAGATG 372
QY 1661 GATCTTCAACCAAGAGCTCACCAGATTTTCCACAGAGATGC 1703
DB 373 GATCTTCAACCAAGAGCTCACCAGATTTTCCACAGAGATGC 415

RESULT 5

US-09-919-580-264
Sequence 264, Application US/09919580
Patent No. US20020110832A1
GENERAL INFORMATION:
APPLICANT: Pyle, Ruth
APPLICANT: Xu, Jiangchun
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.552
CURRENT APPLICATION NUMBER: US/09/919,580
CURRENT FILING DATE: 2001-07-30
NUMBER OF SEQ ID NOS: 934
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 264
LENGTH: 389
TYPE: DNA
ORGANISM: Homo sapiens
US-09-919-580-264

Query Match 5.6%; Score 114; DB 10; Length 389;
Best Local Similarity 100.0%; Pred. No. 1.3e-49;
Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1930 TGCCATCTGTATTAATAAATCCCAATTAAGTCTGTGTAATGTCCTTTTATG 1989
DB 266 TGCCATCTGTATTAATAAATCCCAATTAAGTCTGTGTAATGTCCTTTTATG 325
QY 1990 CTCTTAATTAATTAAGCAGTAATGTCATTTTATGAGATCCCTTAATTAATTAAT 2043
DB 326 CTCTTAATTAATTAAGCAGTAATGTCATTTTATGAGATCCCTTAATTAATTAAT 379

RESULT 6

US-09-927-602-1
Sequence 1, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPLICANT: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
TITLE OF INVENTION: Dystrophy
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2544
TYPE: DNA
ORGANISM: Homo Sapien
FEATURE:
NAME/KEY: CDS
LOCATION: (693)...(1877)
US-09-927-602-1

Query Match 2.9%; Score 59; DB 9; Length 2544;
Best Local Similarity 100.0%; Pred. No. 1.9e-20; Indels 0; Gaps 0;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGGCTCTTACAGCAGCGTGTCTCAAGAGAGTGCCTTCTTCAACCTGCAG 682
DB 1180 AGGCTGCGGCTCTTACAGCAGCGTGTCTCAAGAGAGTGCCTTCTTCAACCTGCAG 1238
RESULT 7
US-09-927-602-38
Sequence 38, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPLICANT: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
TITLE OF INVENTION: Dystrophy
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 38
LENGTH: 48436
TYPE: DNA
ORGANISM: Homo Sapien
US-09-927-602-38

Query Match 2.9%; Score 59; DB 9; Length 48436;
Best Local Similarity 100.0%; Pred. No. 1.9e-20; Indels 0; Gaps 0;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGGCTCTTACAGCAGCGTGTCTCAAGAGAGTGCCTTCTTCAACCTGCAG 682
DB 48024 AGGCTGCGGCTCTTACAGCAGCGTGTCTCAAGAGAGTGCCTTCTTCAACCTGCAG 48082

```
RESULT 8
US-09-816-825-7
; Sequence 7, Application US/09816825
; Patent No. US20010051370A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CON
; CURRENT APPLICATION NUMBER: US/09/816,825
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/045,284
; PRIOR FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-816-825-7

Query Match
Best Local Similarity 1.8%; Score 37; DB 9; Length 37;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 830 AACTCAAGAAGAGGACCAACCTACTATGTGATGC 866
Db 1 AACTCAAGAAGAGGACCAACCTACTATGTGATGC 37

RESULT 9
US-10-007-262-7
; Sequence 7, Application US/10007262
; Publication No. US20020164748A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/10/007,262
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: DNA
; ORGANISM: H. sapiens
US-10-007-262-7

Query Match
Best Local Similarity 1.8%; Score 37; DB 13; Length 37;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 830 AACTCAAGAAGAGGACCAACCTACTATGTGATGC 866
Db 1 AACTCAAGAAGAGGACCAACCTACTATGTGATGC 37

RESULT 10
US-09-867-701-9458
; Sequence 9458, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
US-09-867-701-9458
```

```
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9458
; LENGTH: 167
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9458

Query Match
Best Local Similarity 1.1%; Score 22; DB 10; Length 167;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 23 GAAGACGACAGAGGCTAGAGG 44
Db 3 GAAGACGACAGAGGCTAGAGG 24

RESULT 11
US-09-867-701-9067
; Sequence 9067, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9067
; LENGTH: 171
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9067

Query Match
Best Local Similarity 1.1%; Score 22; DB 10; Length 171;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 23 GAAGACGACAGAGGCTAGAGG 44
Db 9 GAAGACGACAGAGGCTAGAGG 30

RESULT 12
US-09-867-701-9185
; Sequence 9185, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9185
; LENGTH: 187
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9185
```

Query Match 1.1%; Score 22; DB 10; Length 187;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAAGACGACGAGGCTAGAGG 44
DB 32 GAAGACGACGAGGCTAGAGG 53

RESULT 13

US-09-867-701-9284
Sequence 9284 Application US/09867701
Patent No. US2002013237A1
GENERAL INFORMATION:
APPLICANT: Aglate, Paul A.
APPLICANT: Jones, Robert
APPLICANT: Harlocker, Susan L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.497
CURRENT APPLICATION NUMBER: US/09/867,701
CURRENT FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 10912
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9284
LENGTH: 191
TYPE: DNA
ORGANISM: Homo sapien
US-09-867-701-9284

Query Match 1.1%; Score 22; DB 10; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAAGACGACGAGGCTAGAGG 44
DB 29 GAAGACGACGAGGCTAGAGG 50

RESULT 14

US-09-960-352-11099/C
Sequence 11099 Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Mengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathalagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 11099
LENGTH: 398
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 48-BOVMS1-008-Q1-E1-D8
US-09-960-352-11099

Query Match 1.0%; Score 20; DB 10; Length 398;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 GATGGCATTCTGGCTCTAT 194
DB 240 GATGGCATTCTGGCTCTAT 221

RESULT 15

US-10-027-632-140660
Sequence 140660 Application US/10027632

GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 140660
LENGTH: 570
TYPE: DNA
ORGANISM: Human
US-10-027-632-140660

Query Match 1.0%; Score 20; DB 13; Length 570;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 GAAAGACCCCTCCCTCAACC 719
DB 273 GAAAGACCCCTCCCTCAACC 292

Search completed: August 15, 2003, 00:56:03
Job time: 538 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 23:39:27 ; Search time 7280 Seconds

(without alignments)
11480.545 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043

Sequence: 1 gaattccattgtgttgta.....tgggatacctaataaaaaa 2043

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size: 0

Total number of hits satisfying chosen parameters: 4474090

Minimum DB seq length: 50

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_htg_hum:*

31: em_htg_in:*

32: em_htg_inv:*

33: em_htg_other:*

34: em_htg_mus:*

35: em_htg_pln:*

36: em_htg_rtd:*

37: em_htg_man:*

38: em_htg_vtl:*

39: em_sy:*

40: em_htg_hum:*

41: em_htg_mus:*

42: em_htg_rtd:*

43: em_htg_man:*

44: em_sy:*

45: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1722	84.3	1992	9	AF280088	AF280088 Homo sapi
2	1633	80.9	2037	9	BC035282	BC035282 Homo sapi
3	1598	78.2	2011	9	AK026635	AK026635 Homo sapi
4	1597	78.2	183228	9	AC010547	AC010547 Homo sapi
5	1597	78.2	206943	2	AC138848	AC138848 Homo sapi
6	1547	75.7	1979	6	BD127258	BD127258 Homo sapi
7	1547	75.7	1979	6	AK074746	AK074746 Homo sapi
8	1547	75.7	2032	6	AR203335	AR203335 Sequence
9	1547	75.7	2032	6	BD134772	BD134772 Glycosyl
10	1547	75.7	2032	9	AF131235	AF131235 Homo sapi
11	1262	61.8	1333	9	AF149783	AF149783 Homo sapi
12	1153	56.4	2065	6	AR310485	AR310485 Sequence
13	695	34.0	877	6	BD124832	BD124832 primer fo
14	695	34.0	877	6	BD126950	BD126950 primer fo
15	403	19.7	517	6	AX381256	AX381256 Sequence
16	114	5.6	389	6	AX381326	AX381326 Sequence
17	105	5.1	548	6	BD125617	BD125617 primer fo
18	59	2.9	1462	6	AF176839	AF176839 Homo sapi
19	59	2.9	1647	6	AX327330	AX327330 Sequence
20	59	2.9	2170	9	AF176838	AF176838 Homo sapi
21	59	2.9	2544	9	AF219990	AF219990 Homo sapi
22	59	2.9	3278	9	AF246718	AF246718 Homo sapi
23	59	2.9	3786	9	AF280086	AF280086 Homo sapi
24	59	2.9	71503	9	AF219991	AF219991 Homo sapi
25	59	2.9	157358	9	AC009163	AC009163 Homo sapi
26	59	2.9	194832	9	AC025287	AC025287 Homo sapi
27	59	2.9	208185	2	AC009105	AC009105 Homo sapi
28	27	1.3	1926	6	BD134773	BD134773 Glycosyl
29	27	1.3	1926	10	AF131236	AF131236 Mus muscu
30	27	1.3	2201	10	AF109155	AF109155 Mus muscu
31	27	1.3	232745	2	AC130174	AC130174 Rattus no
32	27	1.3	262679	2	AC110314	AC110314 Rattus no
33	27	1.3	288569	2	AC121740	AC121740 Rattus no
34	27	1.2	235173	2	AC133449	AC133449 Rattus no
35	27	1.2	240901	2	AC096328	AC096328 Rattus no
36	27	1.2	242253	2	AC095651	AC095651 Rattus no
37	27	1.2	246877	2	AC137365	AC137365 Rattus no
38	27	1.2	258097	2	AC129764	AC129764 Rattus no
39	27	1.2	271230	2	AC094919	AC094919 Rattus no
40	27	1.1	1740	10	AF176841	AF176841 Mus muscu
41	27	1.1	1989	10	AF176840	AF176840 Mus muscu
42	27	1.1	110000	2	AC099262	AC099262 Continuation (2 of
43	27	1.1	159170	2	AC115914	AC115914 Mus muscu
44	27	1.1	192365	2	AC139891	AC139891 Rattus no
45	27	1.1	194439	2	AC109022	AC109022 Rattus no

ALIGNMENTS

RESULT 1

AF280088

LOCUS

DEFINITION Homo sapiens l-selectin ligand sulfotransferase GST-3 mRNA,

ACCESSION AF280088

VERSION AF280088.1 GI:12060807

KEYWORDS

ORGANISM Homo sapiens (human)

REFERENCE 1 (bases 1 to 1992)

REFERENCE Hemmerich, S., Lee, J.K., Bhakta, S., Bistrup, A., Ruddle, N.R. and Rosen, S.D.

TITLE Chromosomal localization and genomic organization for the
galactose/N-acetylgalactosamine/N-acetylglucosamine
6-O-sulfotransferase gene family
JOURNAL J. Biol. Chem. 274:11815-11822 (1999)
MEDLINE 21096027
PUBMED 1181564
REFERENCE 2 (bases 1 to 1992)
AUTHORS Hemmerich, S., Bhakta, S., Lee, J.-K., Bistrup, A., Ruddie, N.R. and
Rosen, S.D.
TITLE Direct Submision
JOURNAL Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience,
3401 Hillview Ave., Palo Alto, CA 94304, USA
FEATURES
Location/Qualifiers
1..1992
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q23.1-q23.2"
/cell_type="high endothelial"
/tissue_type="tonsil"
1..1992
/gene="GST3"
122..1282
/note="Similar to Homo sapiens sequence encoded by Genbank
Accession Number AF131235"
/product="L-selectin ligand sulfotransferase GST-3"
/codon_start=1
/protein_id="AF131235.1"
/db_xref="GI:12060808"
/translation="MLPKKMLFLVLSOMALILFFHSHNISTLSMKAPERRH
VLVLSWRSSSEVQLFEGHDPVEYLMFPAHVMWTFKSTAMMIMHVRDLIRAVF
LIDMVFDAIMEPGRROSLPOMENRSLCSAPCDLIPDEIIPRACRLCSQP
FEVVEKACRSYSHVYKKEVFNLSLYLKDPSLNIHIVHVRDPRKAEVRSRRTK
GLMIDSRVGMGHOEKLKKEPDYVVOICSOLEIKTIOSLPKAOFERYLYRY
EDLAPAPACTSMYEFVGLLELPHIQTVHNTIRKSGDPAFHTNADALNVSOAM
RMSLEYKYSRLKACGDMNLGTRHNSBOERMLDLSTWTPQIH"
BASE COUNT 482 a 540 c 489 g 481 t
ORIGIN

Query Match 84.3%; Score 1722; DB 9; Length 1992;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1722; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

7 CATTGTTGGGTAGCGAGACAGAGAGGTAGAGAGAAAGCGATGGCCGCTA 66
1 CATTGTTGGGTAGCGAGACAGAGAGGTAGAGAGAAAGCGATGGCCGCTA 60
67 GCAGTGAAGCTCTCAAAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 126
61 GCAGTGAAGCTCTCAAAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
127 AATGCTACTGCTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 186
121 AATGCTACTGCTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
187 GGCCTATATCTTCACATGTACAGCCCAACATCAGCTCCCTGTATAGAGC 246
181 GGCCTATATCTTCACATGTACAGCCCAACATCAGCTCCCTGTATAGAGC 240
247 GCAGGCGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 306
241 GCAGGCGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
307 GCTTTTGGGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366
301 GCTTTTGGGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
367 GACCTTCAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426
361 GACCTTCAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420

427 CGCTCTTCTGTCGACATGAGCGCTCTTATGCTTACATGAACTGTGCCGAGACA 486
421 CGCTCTTCTGTCGACATGAGCGCTCTTATGCTTACATGAACTGTGCCGAGACA 480
487 GTCCAGCCCTTCTGTCGACATGAGCGCTCTTATGCTTACATGAACTGTGCCGAGACA 546
481 GTCCAGCCCTTCTGTCGACATGAGCGCTCTTATGCTTACATGAACTGTGCCGAGACA 540
547 CATCCCAAGATGAAATCATCCCGGGCTCAGTCCAGCTCTGTCAGTCAACAGCC 606
541 CATCCCAAGATGAAATCATCCCGGGCTCAGTCCAGCTCTGTCAGTCAACAGCC 600
607 CTTTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 666
601 CTTTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
667 CTTTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 726
661 CTTTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
727 CGTGCACCTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 786
721 CGTGCACCTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
787 TCTCATGATTGACAGTCCGATTTGATGAGGAGAGAGAGAGAGAGAGAGAGAG 846
781 TCTCATGATTGACAGTCCGATTTGATGAGGAGAGAGAGAGAGAGAGAGAGAG 840
847 CCAGCCCTTCTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 906
841 CCAGCCCTTCTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
907 CCAGTCTTGTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 966
901 CCAGTCTTGTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
967 TCGAGCCCTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1026
961 TCGAGCCCTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
1027 CCATCTTCAAGTCTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1086
1021 CCATCTTCAAGTCTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
1087 CCAGCAGATGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1146
1081 CCAGCAGATGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
1147 TGAAGAGTCTTCAAGTCTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1206
1141 TGAAGAGTCTTCAAGTCTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
1207 CCAGCAGATGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1266
1201 CCAGCAGATGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
1267 TGTCCCTGAGCAATTCATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1326
1261 TGTCCCTGAGCAATTCATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
1327 TCAAGTCAATTCATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1386
1321 TCAAGTCAATTCATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
1387 TCTGAGGATGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1446
1381 TCTGAGGATGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
1447 GTGTCCATGCTTGTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1506
1441 GTGTCCATGCTTGTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
1507 CAGTGAAGAGGATGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1566


```

Db      1501 CAGTGAACAGGAGTATGCTCTTCTTCTTCTGATCTTCCTGCGAGACTTCAG 1560
Qy      1567 AGACTTTGGTGGCTGGAGGCTTATTAAGCAGCAGCATGTAGTGAATGATCCATA 1626
Db      1561 AGACTTTGGTGGCTGGAGGCTTATTAAGCAGCAGCATGTAGTGAATGATCCATA 1620
Qy      1627 ACCTCCCTGTCACATATGCTCCCAATGGGAAATGGATCTTTCACCAAGAAGCTCAGC 1686
Db      1621 ACCTCCCTGTCACATATGCTCCCAATGGGAAATGGATCTTTCACCAAGAAGCTCAGC 1680
Qy      1687 ATTTCCACAGAGATGCGAATCTGAGCCCTTGAGTCCCA 1728
Db      1681 ATTTCCACAGAGATGCGAATCTGAGCCCTTGAGTCCCA 1722

RESULT 2
BC035282
LOCUS
DEFINITION
2037 bp mRNA linear PRI 23-SEP-2002
Homo sapiens, similar to carbohydrate (N-acetylglucosamine 6-O)
sulfotransferase 4, clone MGC:34456 IMAGE:5179826, mRNA, complete
cds.
BC035282
BC035282.1 GI:23273964
MGC.
Homo sapiens (human)
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 2037)
Strausberg, R.
Direct Submission
Submitted (31-JUL-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2530,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: mgc-help@nci.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunnarone, P.H., Garcia, A.M., Lu, X., Huiyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: IRAC Plate: 50 Row: a Column: 7
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA g1: 5031734.
Location/Qualifiers
1. 2037
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:34456 IMAGE:5179826"
/tissue_type="Brain, Lung, Testis, adult, pooled whole"
/clone_lib="NIH_MGC_115"
/lab_host="DH10B"
/vec="Vector: PCMV-SPORT6"
/size="126..1238"
/codon_start=1
/product="Similar to carbohydrate (N-acetylglucosamine
6-O) sulfotransferase 4"
/protein_id="AAH35282.1"
/db_xref="GI:23273965"
/translation="MAILAFHMYSHNISLSMKAQPERHVLVLSMWSGSSFWGQ"
FEATURES
source

```

```

LFGQHPDVFYIEMEPAMHVMWTFKOSTAMLMHMAVRDLIRAVFLCDKNSVPDAMEPGR
KRSLEFQWENSRLALCSAPACDIIIPDEITIRAHORLICSOOPFEVYKACRSYHVL
KVFREPLQSLYPLKDPASLNHIVHVRPDRVRSRERTKIDIMDSITVNGOHO
KIKKEDPITYMOVITCOSOLEITKTITSLKALQERILVREDIARAPVQIQRME
FVGLIEPLQTVNHNITRGKMGDHFHFNADALVNSQAMRWISLPERVSRLOKAC
GDAMNLGYSVESEQDRNLDLISTVTPVPROIH"
BASE COUNT      507 a      546 c      481 g      503 t
ORIGIN
Query Match      80.9%; Score 1653; DB 9; Length 2037;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1653; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      51 GCGCATGCCCGGCTAGCAGTACGCTCTCTAAAAGCAGCAGGAGCCCAAGCCAAAGG 110
Db      1 GCGCATGCCCGGCTAGCAGTACGCTCTCTAAAAGCAGCAGGAGCCCAAGCCAAAGG 60
Qy      111 TCTTCACATTGACGACAAATGCTACTGCTTAAAAAATGAAGCTCTGTTCTGTTT 170
Db      61 TCTTCACATTGACGACAAATGCTACTGCTTAAAAAATGAAGCTCTGTTCTGTTT 120
Qy      171 CCCAGATGGCCATTTTGGCTCTATTCTTCCACATGTACAGCCAAACATCAGCTCCCTGT 230
Db      121 CCCAGATGGCCATTTTGGCTCTATTCTTCCACATGTACAGCCAAACATCAGCTCCCTGT 180
Qy      231 CTATGAAGGACACAGCCCGAGCGCATGCTGCTGTCTGTCTCTGCGGCTGTGGCT 290
Db      181 CTATGAAGGACACAGCCCGAGCGCATGCTGCTGTCTGTCTCTGCGGCTGTGGCT 240
Qy      291 CTCTCTTTTGTGGGGACCTTTTGGGAGCACCAGATGTTTCTTACCTGATGAGCCCG 350
Db      241 CTCTCTTTTGTGGGGACCTTTTGGGAGCACCAGATGTTTCTTACCTGATGAGCCCG 300
Qy      351 CTTGGCAGCTGTGATATACCTTTCAAGCAGACAGCCGCTGGATCTGCATAGGCTGTGC 410
Db      301 CTTGGCAGCTGTGATATACCTTTCAAGCAGACAGCCGCTGGATCTGCATAGGCTGTGC 360
Qy      411 GGGATCTGATACGGGCGGCTCTTCTTGGCGATGAGCGTCTTGAATGCTTATGGAAC 470
Db      361 GGGATCTGATACGGGCGGCTCTTCTTGGCGATGAGCGTCTTGAATGCTTATGGAAC 420
Qy      471 CTGGTCCCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAAACGCCGCGCTGTGTTCTG 530
Db      421 CTGGTCCCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAAACGCCGCGCTGTGTTCTG 480
Qy      531 CACCTGCTGTGATATATCCACAGATGAATATCCCGGGGCTCAGTGCAGGCGTCC 590
Db      481 CACCTGCTGTGATATATCCACAGATGAATATCCCGGGGCTCAGTGCAGGCGTCC 540
Qy      591 TGTGCACTCAACAGCCCTTGGAGGTGGTGGAGAGGCGCTGCGCTCTACAGCCACGTGG 650
Db      541 TGTGCACTCAACAGCCCTTGGAGGTGGTGGAGAGGCGCTGCGCTCTACAGCCACGTGG 600
Qy      651 TGTCAAGAGAGGTGCGCTTCTTCAACCTGAGTCCCTTACCCGCTCTGAAGACCCCT 710
Db      601 TGTCAAGAGAGGTGCGCTTCTTCAACCTGAGTCCCTTACCCGCTCTGAAGACCCCT 660
Qy      711 CCTCAACCTGCAATATGTGCACTGCTCGGGAGACCCCGGGCGGTGTTCGTTCCGGAG 770
Db      661 CCTCAACCTGCAATATGTGCACTGCTCGGGAGACCCCGGGCGGTGTTCGTTCCGGAG 720
Qy      771 AAGCACAAGAGGAGATCTCATGTTTACAGTGCATTTGATGGGGGACATGAGCAAA 830
Db      721 AAGCACAAGAGGAGATCTCATGTTTACAGTGCATTTGATGGGGGACATGAGCAAA 780
Qy      831 AACTCAAGAGAGGAGACCAACCTTACTATGTGAGTCAAGTCAATCTGCCAAGCAGCTGG 890
Db      781 AACTCAAGAGAGGAGACCAACCTTACTATGTGAGTCAAGTCAATCTGCCAAGCAGCTGG 840
Qy      891 AGATCTCAAGAGACCATCCTCTTGGCCCAAGGCGCTGCAAGAACGCTTACCTGCTGTGC 950
Db      841 AGATCTCAAGAGACCATCCTCTTGGCCCAAGGCGCTGCAAGAACGCTTACCTGCTGTGC 900

```

```

OY 951 GCTATGAGAGCTGGCTGAGCCCTGCTGGCCAGACTTCCGAAATGATGATTCGTG 1010
Db 901 GCTATGAGAGCTGGCTGAGCCCTGCTGGCCAGACTTCCGAAATGATGATTCGTG 960
OY 1011 GATTGGAATCTTGGCCCATCTTTCAGACCTGGGTGATTAACATCACCAGGCAAGGCA 1070
Db 961 GATTGGAATCTTGGCCCATCTTTCAGACCTGGGTGATTAACATCACCAGGCAAGGCA 1020
OY 1071 TGGGTGACACGCTTTCACACAAATGCCAGGATGCCCTTATGTCCTCCAGGCTTGGC 1130
Db 1021 TGGGTGACACGCTTTCACACAAATGCCAGGATGCCCTTATGTCCTCCAGGCTTGGC 1080
OY 1131 GCTGCTCTTGGCCCATATGAAAAGTTTCTGCACTTCAAGACCTGGTGGCATGCA 1190
Db 1081 GCTGCTCTTGGCCCATATGAAAAGTTTCTGCACTTCAAGACCTGGTGGCATGCA 1140
OY 1191 ATTGCTGGGCTACCGCCACGCTGATCTGAAACAGAAAGAAACCTTGTGTCGATC 1250
Db 1141 ATTGCTGGGCTACCGCCACGCTGATCTGAAACAGAAAGAAACCTTGTGTCGATC 1200
OY 1251 TTCTGCTACCTGAGACTGCTCCCTGAGCAATCCACTAAGAGGTTGAGAAGCTTGTG 1310
Db 1201 TTCTGCTACCTGAGACTGCTCCCTGAGCAATCCACTAAGAGGTTGAGAAGCTTGTG 1260
OY 1311 CCACCTGGTGTACGCTCAGTCACTTCTCTGAAAGCTTCTGACCTTGCCTACATCT 1370
Db 1261 CCACCTGGTGTACGCTCAGTCACTTCTCTGAAAGCTTCTGACCTTGCCTACATCT 1320
OY 1371 GAGCCTTAACATGCTGTGTGGTATCACTGATGATGAGTGTGTCACACGCTGCTC 1430
Db 1321 GAGCCTTAACATGCTGTGTGGTATCACTGATGATGAGTGTGTCACACGCTGCTC 1380
OY 1431 AAGCAGAAAGACTTTTGTGTCATGCTTGTCTAGAAAACAGACTGGGGAACCTTATGT 1490
Db 1381 AAGCAGAAAGACTTTTGTGTCATGCTTGTCTAGAAAACAGACTGGGGAACCTTATGT 1440
OY 1491 GAGCAGCAATCCCAACCAATGAAACAGGATATGCTCTCTCTTCTTGTGATCTTCTG 1550
Db 1441 GAGCAGCAATCCCAACCAATGAAACAGGATATGCTCTCTCTTCTTGTGATCTTCTG 1500
OY 1551 TCTGGGCACTTTCAGAGACTTGTGGGCTGAGAGGCTTATTAAGCAGCAGAGATCTAG 1610
Db 1501 TCTGGGCACTTTCAGAGACTTGTGGGCTGAGAGGCTTATTAAGCAGCAGAGATCTAG 1560
OY 1611 TGGATTTGATCAATAAACCCTCTGTCACATCTTGGCCAAATGGGGAATGATCTTTCAC 1670
Db 1561 TGGATTTGATCAATAAACCCTCTGTCACATCTTGGCCAAATGGGGAATGATCTTTCAC 1620
OY 1671 CAAGAGCTCAGCAGATTTTCCACAGAGATGC 1703
Db 1621 CAAGAGCTCAGCAGATTTTCCACAGAGATGC 1653

```

```

RESULT 3
AK026635
LOCUS
DEFINITION
2011 bp mRNA linear prt 29-sep-2000
Home sapiens cDNA: FLJ22982 fis, clone KAT11454, highly similar to
AF131235 Homo sapiens N-acetylglucosamine 6-O-sulfotransferase
mRNA.
ACCESSION
AK026635
VERSION
AK026635.1 GI:10439531
KEYWORDS
oligo capping, fis (full insert sequence).
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS
Matenabe, K., Kumagai, A., Itakura, S., Yamazaki, M., Tashiro, H.,
Ota, T., Suzuki, Y., Odayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
Nakamura, Y., Isogai, T. and Sugano, S.
NEBO human cDNA sequencing project
Unpublished
TITLE
NEBO human cDNA sequencing project
JOURNAL
REFERENCE
2 (bases 1 to 2011)

```

```

AUTHORS
Sugano, S., Suzuki, Y., Ota, T., Odayashi, M., Nishi, T., Isogai, T.,
Shibahara, T., Tanaka, T. and Nakamura, Y.
DIRECT SUBMISSION
TITLE
Sugano, S., Suzuki, Y., Ota, T., Odayashi, M., Nishi, T., Shibahara, T., Tanaka, T. and Nakamura, Y.
JOURNAL
Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
Japan (E-mail:cdna@leims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT
NEBO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing: Research Association for Biotechnology: cDNA library
construction, 5'-6' and 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source
1..2011
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="KAT11454"
/cell_line="KATO III"
/cell_type="signet-ring cell carcinoma"
/clone_1ib="KAT"
/Note="Cloning vector pME18SFL3"
misc_feature
1..2011
/Note="highly similar to AF131235 Homo sapiens
N-acetylglucosamine 6-O-sulfotransferase mRNA"
BASE COUNT
482 a 541 c 475 g 513 t
ORIGIN
Query Match 78.2%; Score 1598; DB 9; Length 2011;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1598; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 106 CAAGCTTCCACTTACGACAAATGCTACTGCTTAAAGAAAGCTCTCTCTTCT 165
Db 50 CAAGCTTCCACTTACGACAAATGCTACTGCTTAAAGAAAGCTCTCTCTCTTCT 109
OY 166 GGTTCCTCCAGATGGCCATCTTGGCTTATCTTCCACATGATACAGCACAATCAGCTC 225
Db 110 GGTTCCTCCAGATGGCCATCTTGGCTTATCTTCCACATGATACAGCACAATCAGCTC 169
OY 226 CCTGTATGAAGGACACACCCGAGCGCATGCACTGCTGTCTCTCTCTCTCTCTCTCT 285
Db 170 CCTGTATGAAGGACACACCCGAGCGCATGCACTGCTGTCTCTCTCTCTCTCTCTCT 229
OY 286 TGGCTCTCTTGTGTGGGAGAGCTTTTGGGAGACACCGAGATGTTTCTACCTGATGA 345
Db 230 TGGCTCTCTTGTGTGGGAGAGCTTTTGGGAGACACCGAGATGTTTCTACCTGATGA 289
OY 346 GCCCGCTGGCAGCTGTGATGACCTTCAAGCAGACACCGCTGTGATGCTCAGATGCG 405
Db 290 GCCCGCTGGCAGCTGTGATGACCTTCAAGCAGACACCGCTGTGATGCTCAGATGCG 349
OY 406 TGTGCGGATCTGATACGCGGCTCTTCTTGTGCGACATGAGCGCTTTGATGCTCACT 465
Db 350 TGTGCGGATCTGATACGCGGCTCTTCTTGTGCGACATGAGCGCTTTGATGCTCACT 409
OY 466 GGAACCTGTGTCCTGGGAGAGAGTCCAGCTCTTCAAGTGGGAGAAACAGCGGGGCTGTG 525
Db 410 GGAACCTGTGTCCTGGGAGAGAGTCCAGCTCTTCAAGTGGGAGAAACAGCGGGGCTGTG 469
OY 526 TTCTGCACTGCTGTGATCATCCACAGATGAATCATCCCGGGGCTCACTGACAG 585
Db 470 TTCTGCACTGCTGTGATCATCCACAGATGAATCATCCCGGGGCTCACTGACAG 529
OY 586 GCTCTGTGCACTGAACACCCCTTTGAGGTGTGGAAGAGCGCTGCTCTCTACAGCCA 645
Db 530 GCTCTGTGCACTGAACACCCCTTTGAGGTGTGGAAGAGCGCTGCTCTCTACAGCCA 589
OY 646 CGTGTGCTCAAGAGAGGTGCTCTTCAACCTGCACTGCTCTCTCTCTCTCTCTCTCT 705

```

Db 590 CGTGTGCTCAGAGAGTGGCTTCTTCAACCTGCACTGCCCTCTACCCGCTGTGAAGA 649
 QY 706 CCCCTCCCTCAACCTGCATATCGACACGTGTCGGGACCCCGGGCGGTTCGCTTC 765
 Db 650 CCCCTCCCTCAACCTGCATATCGACACGTGTCGGGACCCCGGGCGGTTCGCTTC 709
 QY 766 CCGAGAACGACAAAGGAGATCTCATGATGTGACAGTGCATTTGTGAGGGGACGATGA 825
 Db 710 CCGAGAACGACAAAGGAGATCTCATGATGTGACAGTGCATTTGTGAGGGGACGATGA 769
 QY 826 GCAAAAACCTCAAGAGAGAGACCAACCTCTATGTATGAGTCTATCTGCCAAACCA 885
 Db 770 GCAAAAACCTCAAGAGAGAGACCAACCTCTATGTATGAGTCTATCTGCCAAACCA 829
 QY 886 GCTGAGATCTACAAAGACATCCAGTCTTGGCCCAAGGCCCTGAGAGAGCTACCTGCT 945
 Db 830 GCTGAGATCTACAAAGACATCCAGTCTTGGCCCAAGGCCCTGAGAGAGCTACCTGCT 889
 QY 946 TGTGCGCTATGAGAGACCTGGCTGAGGCCCTGTGGCCAGACTTCCGAATGTATGAAT 1005
 Db 890 TGTGCGCTATGAGAGACCTGGCTGAGGCCCTGTGGCCAGACTTCCGAATGTATGAAT 949
 QY 1006 CGTGGGATTTGGAATTTCTTGGCCCATCTTCAAGACCTGGGTGATTAATCAACCCGAGCAA 1065
 Db 950 CGTGGGATTTGGAATTTCTTGGCCCATCTTCAAGACCTGGGTGATTAATCAACCCGAGCAA 1009
 QY 1066 GGGCATGGGTGACACAGCTTTCACACAAATGCGAGGATGCCCTTAATCTTCCCAAGC 1125
 Db 1010 GGGCATGGGTGACACAGCTTTCACACAAATGCGAGGATGCCCTTAATCTTCCCAAGC 1069
 QY 1126 TTGGCGCTGTCTTGGCCCATATGAAAAGTTTCTCGACTTCAGAAACCTGTGGCATTC 1185
 Db 1070 TTGGCGCTGTCTTGGCCCATATGAAAAGTTTCTCGACTTCAGAAACCTGTGGCATTC 1129
 QY 1186 CATGAATTTCTGTGGCTACCCGCACTCAATCTGAACAAAGAAACCTGTGCT 1245
 Db 1130 CATGAATTTCTGTGGCTACCCGCACTCAATCTGAACAAAGAAACCTGTGCT 1189
 QY 1246 GGAATCTTCTGTACCTGAGTGCCTGAGCAAAATCCACTAAGAGGTTGAGAAGCTT 1305
 Db 1190 GGAATCTTCTGTACCTGAGTGCCTGAGCAAAATCCACTAAGAGGTTGAGAAGCTT 1249
 QY 1306 TGTGCGACCTGTGTGTCAGCTCACTCTTCTGAAATCTTCTGAGCCTTGCCTACA 1365
 Db 1250 TGTGCGACCTGTGTGTCAGCTCACTCTTCTGAAATCTTCTGAGCCTTGCCTACA 1309
 QY 1366 TCTCTGAGCTTAACTACATGCTGTGTGATACACTGATGAGTGTGTCACACG 1425
 Db 1310 TCTCTGAGCTTAACTACATGCTGTGTGATACACTGATGAGTGTGTCACACG 1369
 QY 1426 TGTCAAGCAAGAGAGCTTGTGTGCATGCTGTGTCTGAAAGCAAGACTGGGAACT 1485
 Db 1370 TGTCAAGCAAGAGAGCTTGTGTGCATGCTGTGTCTGAAAGCAAGACTGGGAACT 1429
 QY 1486 TATGTAGAGCAACATCCACACAGTGAACAGGGTATTCCTTCTTCTTCTGTATCT 1545
 Db 1430 TATGTAGAGCAACATCCACACAGTGAACAGGGTATTCCTTCTTCTTCTGTATCT 1489
 QY 1546 TCTGTCTGTGGGAGACTTCAAGAGACTTGTGGCCCTGAGGCTTATTAAGCAGACAGT 1605
 Db 1490 TCTGTCTGTGGGAGACTTCAAGAGACTTGTGGCCCTGAGGCTTATTAAGCAGACAGT 1549
 QY 1606 ATGAGTGAATTAATCATTAACCTCCCTCTTCCACATCTTGCCCAATGGGGAATGATCT 1665
 Db 1550 ATGAGTGAATTAATCATTAACCTCCCTCTTCCACATCTTGCCCAATGGGGAATGATCT 1609
 QY 1666 TTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 1703
 Db 1610 TTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 1647

RESULT 4
 AC010547

LOCUS AC010547 183228 bp DNA linear PRI 29-SEP-2001
 DEFINITION Homo sapiens chromosome 16 clone RP11-510M2, complete sequence.
 AC010547
 VERSION AC010547.9 GI:15808510
 KEYWORDS htc.
 SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 183228)
 DOE Joint Genome Institute.
 Sequencing of Human Chromosome 16
 Unpublished
 2 (bases 1 to 183228)
 DOE Joint Genome Institute.
 Direct Submission
 Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 3 (bases 1 to 183228)
 DOE Joint Genome Institute.
 Direct Submission
 Submitted (29-SEP-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 On Sep 29, 2001 this sequence version replaced gi:1589436.
 COMMENT
 Sequence Quality Assessment:
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

 Sequence Quality Assessment:
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

 FEATURES
 source location/Qualifiers
 1..183228
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-510M2"
 BASE COUNT 51901 a 42098 c 40026 g 49203 t
 ORIGIN
 Query Match 78.2%; Score 1597; DB 9; Length 183228;
 Best local Similarity 100.0%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 107 AAGTCTTCACCTGACGACAAATGCTACTGCTTAAATAAATGAAGCTGCTGCTTCTG 166
 Db 142584 AAGTCTTCACCTGACGACAAATGCTACTGCTTAAATAAATGAAGCTGCTGCTTCTG 142643
 QY 167 GTTTCACAGATGGGACATCTGGCTCTATTTCTTCACATGTACAGCCCAACATCAGCTCC 226
 Db 142644 GTTTCACAGATGGGACATCTGGCTCTATTTCTTCACATGTACAGCCCAACATCAGCTCC 142703
 QY 227 CTGTCTATGAGAGCAGCCGAGCGCATGCTGCTGCTGCTTCTTCTGCGCTCT 286
 Db 142704 CTGTCTATGAGAGCAGCCGAGCGCATGCTGCTGCTGCTTCTTCTGCGCTCT 142763
 QY 287 GGCCTCTCTTTGTGTGGGAGACTTTTGGAGCAGCACCAGAGTCTTCTACCTGATGAG 346
 Db 142764 GGCCTCTCTTTGTGTGGGAGACTTTTGGAGCAGCACCAGAGTCTTCTACCTGATGAG 142823

```

OY 347 CCGGCTGGACGCTGTGATGACCTTCAAGCAGACGACCGCTGTGATGCTGACATGCGT 406
DB 142824 CCGGCTGGACGCTGTGATGACCTTCAAGCAGACGACCGCTGTGATGCTGACATGCGT 142883
OY 407 GTGCGGGATCTATACGGGGCTCTTCTTGTGCGACATGAGCGTCTTGTATGCTTACATG 466
DB 142884 GTGCGGGATCTATACGGGGCTCTTCTTGTGCGACATGAGCGTCTTGTATGCTTACATG 142943
OY 467 GAACCTGGTCCCGGAGACAGTCCAGCCCTTTCAGTGGAGAACAGCGCGGGCCGTGT 526
DB 142944 GAACCTGGTCCCGGAGACAGTCCAGCCCTTTCAGTGGAGAACAGCGCGGGCCGTGT 143003
OY 527 TCTGCACCTGCTGTGATCATATCCACAAGATGAATCATCCCGGGCTCATCTGACAG 586
DB 143004 TCTGCACCTGCTGTGATCATATCCACAAGATGAATCATCCCGGGCTCATCTGACAG 143063
OY 587 CTCTGTGAGTCAACAGCCCTTTCAGTGGAGAACAGCGCGGGCTCTTACAGCCAC 646
DB 143064 CTCTGTGAGTCAACAGCCCTTTCAGTGGAGAACAGCGCGGGCTCTTACAGCCAC 143123
OY 647 GTGCTGTCAAGAGAGTGGCTTCTTCAACCTGACGTCCTCTACCCGCTGCTGAAGAC 706
DB 143124 GTGCTGTCAAGAGAGTGGCTTCTTCAACCTGACGTCCTCTACCCGCTGCTGAAGAC 143183
OY 707 CCTCTCTCAACCTGATATGTGTGACCTGTCCGGGACCCCGGGCCGTGTCCGTTCC 766
DB 143184 CCTCTCTCAACCTGATATGTGTGACCTGTCCGGGACCCCGGGCCGTGTCCGTTCC 143243
OY 767 CGAGAACGCAAAAGGAGATCTCATGATTGACAGTGTGATGGGGAGCATGAG 826
DB 143244 CGAGAACGCAAAAGGAGATCTCATGATTGACAGTGTGATGGGGAGCATGAG 143303
OY 827 CAAAACCTCAAGAGAGACCAACCTCATATGTGATGATGATGATGATGATGATGATG 886
DB 143304 CAAAACCTCAAGAGAGACCAACCTCATATGTGATGATGATGATGATGATGATGATG 143363
OY 887 CTGGAGATCTTCAAGACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 946
DB 143364 CTGGAGATCTTCAAGACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 143423
OY 947 GTGCGCTATGAGACCTGCTGAGACCCCTGTGCGCCAGTCTCCGATGATGATGATG 1006
DB 143424 GTGCGCTATGAGACCTGCTGAGACCCCTGTGCGCCAGTCTCCGATGATGATGATG 143483
OY 1007 GTGGAGTTGGAATTTCTGCCCCATCTTCAAGACCTGCTGCTGCTGCTGCTGCTGCT 1066
DB 143484 GTGGAGTTGGAATTTCTGCCCCATCTTCAAGACCTGCTGCTGCTGCTGCTGCTGCT 143543
OY 1067 GGCATGGGTGACCGCTTTCACACAAATGCGAGGATGCTTAAATGCTTCCAGGCT 1126
DB 143544 GGCATGGGTGACCGCTTTCACACAAATGCGAGGATGCTTAAATGCTTCCAGGCT 143603
OY 1127 TGGCGCTGCTTCTTCCCTATGAAAAAGTTTCTGCACTTCAAGAAAGCTGTGGCGATG 1186
DB 143604 TGGCGCTGCTTCTTCCCTATGAAAAAGTTTCTGCACTTCAAGAAAGCTGTGGCGATG 143663
OY 1187 ATGAATTTGCTGGCTACCGCAGCTGATGATGAACAAGAACAGAAACCTGTTGCTG 1246
DB 143664 ATGAATTTGCTGGCTACCGCAGCTGATGATGAACAAGAACAGAAACCTGTTGCTG 143723
OY 1247 GATCTTGTCTACCTGAGCTGCTTCCAGCAATCCATAAGAGGTTGAGAAAGCTTT 1306
DB 143724 GATCTTGTCTACCTGAGCTGCTTCCAGCAATCCATAAGAGGTTGAGAAAGCTTT 143783
OY 1307 GGTGCGACCTGCTGATGAGCTGATGATGATGATGATGATGATGATGATGATGATG 1366
DB 143784 GGTGCGACCTGCTGATGAGCTGATGATGATGATGATGATGATGATGATGATGATG 143843
OY 1367 CTCTGAGCCTTAACTACATGCTGTGGGATATCACATGATGATGATGATGATGATG 1426
DB 143844 CTCTGAGCCTTAACTACATGCTGTGGGATATCACATGATGATGATGATGATGATG 143903

```

```

OY 1427 GCTCAAGCAGAGGACTTTTGTGTCCATGCTGTGTCTAGAAACAGACTGGGAACTT 1486
DB 143904 GCTCAAGCAGAGGACTTTTGTGTCCATGCTGTGTCTAGAAACAGACTGGGAACTT 143963
OY 1487 ATGTGAGCAGCATCCACACAGTGAACAGGGATATGCTCTTCTTCTTCTTCTTCTT 1546
DB 143964 ATGTGAGCAGCATCCACACAGTGAACAGGGATATGCTCTTCTTCTTCTTCTTCTT 144023
OY 1547 CTTGTCTGGGAGACTTCAAGACTTTTGGCGCTGAGAGCTTATTAAGCAGCAGACAGTA 1606
DB 144024 CTTGTCTGGGAGACTTCAAGACTTTTGGCGCTGAGAGCTTATTAAGCAGCAGACAGTA 144083
OY 1607 TCAGTGAATTAATCCATTAACCTCCCTGTGTCACATCTTCCCAATGGGAAATGATCTT 1666
DB 144084 TCAGTGAATTAATCCATTAACCTCCCTGTGTCACATCTTCCCAATGGGAAATGATCTT 144143
OY 1667 TCACCAAGAGCTCCACCATTTTCCACAGAGATC 1703
DB 144144 TCACCAAGAGCTCCACCATTTTCCACAGAGATC 144180

RESULT 5
AC138848/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT

AC138848 206943 bp DNA linear HTG 21-JAN-2003
Homo sapiens chromosome 5 clone RP11-1301B21, WORKING DRAFT
SEQUENCE, 7 unordered pieces.
AC138848
AC138848.1 GI:27805260
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 206943)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 206943)
DOE Joint Genome Institute.
Direct Submission
Submitted (21-JAN-2003) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 2719263
Center clone name: RCT1-11_1301B21
-----
Summary Statistics
Consensus quality: 202039 bases at least Q40
Consensus quality: 202804 bases at least Q30
Consensus quality: 203337 bases at least Q20
Estimated insert size: 175000; agarose-fp estimation
Estimated insert size: 206343; sum-of-ctdigs estimation
Quality coverage: 15.28 in Q20 bases; agarose-fp estimation
Quality coverage: 12.96 in Q20 bases; sum-of-ctdigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1219: contig of 1219 bp in length
* 1220 1319: gap of unknown length
* 1320 2493: contig of 1174 bp in length
* 2494 2593: gap of unknown length
* 2594 3666: contig of 1073 bp in length
* 3667 3766: gap of unknown length

```

```

*      3767      5167: contig of 1401 bp in length
*      5168      5267: gap of unknown length
*      5268      12818: contig of 7551 bp in length
*      12918      12919: gap of unknown length
*      36168      36169: contig of 23251 bp in length
*      36170      36269: gap of unknown length
*      36270      206943: contig of 170674 bp in length.
FEATURES
    source
        1. 206943
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone="RP11-1301B21"
            /clone_1db="RP11 human BAC library 11"
BASE COUNT      56312 a 44122 c 46163 g 59225 t 1121 others
ORIGIN
Query Match      78.2%; Score 1597; DB 2; Length 206943;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*      107      AAGCTCTCCACTTCAGACACAGCTACTGCTTAAATAATGAAGCTCTGCTGTTCTG 166
*      125368      AAGCTCTCCACTTCAGACACAGCTACTGCTTAAATAATGAAGCTCTGCTGTTCTG 125309
*      167      GTTCCAGATGGCATCTGGCTCTATTCTTCACATGTACAGCCACAACATAGCTCC 226
*      125308      GTTCCAGATGGCATCTGGCTCTATTCTTCACATGTACAGCCACAACATAGCTCC 125249
*      227      CTGTCTATGAAGGACACAGCCGAGCGCATGACAGCTGTGTTCTGTTCTCTGCGCTCT 286
*      125248      CTGTCTATGAAGGACACAGCCGAGCGCATGACAGCTGTGTTCTCTGCGCTCT 125189
*      287      GGCCTCTCTTTTGTGGGGACGCTTTTGGGACAGCACCAGATGTTTCTACCTGATGAG 346
*      125188      GGCCTCTCTTTTGTGGGGACGCTTTTGGGACAGCACCAGATGTTTCTACCTGATGAG 125129
*      347      CCCGCTGGACGCTGTGGATGACCTTCAAGACAGACACCGCTGGATGCTGACATGGCT 406
*      125128      CCCGCTGGACGCTGTGGATGACCTTCAAGACAGACACCGCTGGATGCTGACATGGCT 125069
*      407      GTGGGGATCTGATACGGGCGCTCTTGTGCGACATGAGCGCTTTGATGCTTACATG 466
*      125068      GTGGGGATCTGATACGGGCGCTCTTGTGCGACATGAGCGCTTTGATGCTTACATG 125009
*      467      GAACCTGCTCCCGGAGACAGATCCAGCTCTTTCAGTGGGAGAACAGCCGGCCCTGTGT 526
*      125008      GAACCTGCTCCCGGAGACAGATCCAGCTCTTTCAGTGGGAGAACAGCCGGCCCTGTGT 124949
*      527      TCTGCACCTGCTGCTGATCATCTCCACAAGATGAATTCATCCCGGGGCTCACTGGAG 586
*      124948      TCTGCACCTGCTGCTGATCATCTCCACAAGATGAATTCATCCCGGGGCTCACTGGAG 124889
*      587      CTCCTGTGAGTCAACAGCCCTTTGAGGTGTGGAGAGGCGCTCCCTCTACAGCCAC 646
*      124888      CTCCTGTGAGTCAACAGCCCTTTGAGGTGTGGAGAGGCGCTCCCTCTACAGCCAC 124829
*      647      GTGGTGTCAAGAGAGGCGCTTCTTCAACCTGCAATGCCCTCTACCCGCTGCTAAAGAC 706
*      124828      GTGGTGTCAAGAGAGGCGCTTCTTCAACCTGCAATGCCCTCTACCCGCTGCTAAAGAC 124769
*      707      CCTCCCTCAACCTGCAATGCTGTCACCTGTGCGGAGACCCCGGGCGCTTCCGTTCC 766
*      124768      CCTCCCTCAACCTGCAATGCTGTCACCTGTGCGGAGACCCCGGGCGCTTCCGTTCC 124709
*      767      CGAGACGCAAAAGGAGATCTCATATTGACAGTGCATTTGATGGGGCAGCATGAG 826
*      124708      CGAGACGCAAAAGGAGATCTCATATTGACAGTGCATTTGATGGGGCAGCATGAG 124649
*      827      CAAAACCTCAAGAGAGAGCAACCTACTATGTGATGCAAGTGCATCTGCCAAAGCCAG 886
*      124648      CAAAACCTCAAGAGAGAGCAACCTACTATGTGATGCAAGTGCATCTGCCAAAGCCAG 124589

```

```

OY      887      CTGAGATCTTACAGACCATCCAGCTCTTGCCCAAGGCCCTGCAGAAACGCTACTGCTT 946
DB      124588      CTGAGATCTTACAGACCATCCAGCTCTTGCCCAAGGCCCTGCAGAAACGCTACTGCTT 124529
OY      947      GTGGCTATGAGAGCATGCTGCTGAGCCCTGTGGCCAGACATTCGCGAATGATGAATTC 1006
DB      124528      GTGGCTATGAGAGCATGCTGCTGAGCCCTGTGGCCAGACATTCGCGAATGATGAATTC 124469
OY      1007      GTGGATTTGCAATTTCTTGCCCATCTTACAGCTTGGTGTGCTTAACATCACCAGGAGC 1086
DB      124468      GTGGATTTGCAATTTCTTGCCCATCTTACAGCTTGGTGTGCTTAACATCACCAGGAGC 124409
OY      1067      GGCATGGGTGACACAGGCTTTCACACAAATGCGAGGATGCGCTTAATGTTCCAGGCT 1126
DB      124408      GGCATGGGTGACACAGGCTTTCACACAAATGCGAGGATGCGCTTAATGTTCCAGGCT 124349
OY      1127      TGCGCTGTGCTTGTGCTTATGAAAAGTTCCTGACTTCAGAAAAGCTGTGGCGATGCC 1186
DB      124348      TGCGCTGTGCTTGTGCTTATGAAAAGTTCCTGACTTCAGAAAAGCTGTGGCGATGCC 124289
OY      1187      ATGAATTTGCTGGGCTACCGCCAGCTCAGATCTGAACAGACAGAAACCTTTGCTG 1246
DB      124288      ATGAATTTGCTGGGCTACCGCCAGCTCAGATCTGAACAGACAGAAACCTTTGCTG 124229
OY      1247      GATCTTGTCTACTCTGAGCTGTCCCTGAGCAATTCACCTAAGAGGTTGAGAAAGCTTT 1306
DB      124228      GATCTTGTCTACTCTGAGCTGTCCCTGAGCAATTCACCTAAGAGGTTGAGAAAGCTTT 124169
OY      1307      GCTGCCACCTGTGCTAGCCCTCAGTCACTTCTCTGATGCTTCTGAGCTTGCCTCAT 1366
DB      124168      GCTGCCACCTGTGCTAGCCCTCAGTCACTTCTCTGATGCTTCTGAGCTTGCCTCAT 124109
OY      1367      CTCTGAGCTTAACTAATGCTGTGTGGTATCACACTGAGTGTGATGTTGTCCACACT 1426
DB      124108      CTCTGAGCTTAACTAATGCTGTGTGGTATCACACTGAGTGTGATGTTGTCCACACT 124049
OY      1427      GCTCAAGACAGAAAGACTTTTGTGTCATGCTTGTGTAGAAAACAGACTGGGAACTT 1486
DB      124048      GCTCAAGACAGAAAGACTTTTGTGTCATGCTTGTGTGTAGAAAACAGACTGGGAACTT 123989
OY      1487      ATGTGACGACGACATCCACAGTGAAGAGGATGATGCTTCTTCTTCTTCTGATCTT 1546
DB      123988      ATGTGACGACGACATCCACAGTGAAGAGGATGATGCTTCTTCTTCTTCTGATCTT 123929
OY      1547      CCTGTCTGGGACAGACTTCAAGAGACTTGTGGCTGGAGGCTTATTAAGCACAGACATA 1606
DB      123928      CCTGTCTGGGACAGACTTCAAGAGACTTGTGGCTGGAGGCTTATTAAGCACAGACATA 123869
OY      1607      TCAAGTGAATTAATGATCAATTAACCTCCCTGTTCACATCTTGGCCAAATGGGAATGATCTT 1666
DB      123868      TCAAGTGAATTAATGATCAATTAACCTCCCTGTTCACATCTTGGCCAAATGGGAATGATCTT 123809
OY      1667      TCACCAAAAGAGCTCAGCAGCATTTTTCACAGAGATGC 1703
DB      123808      TCACCAAAAGAGCTCAGCAGCATTTTTCACAGAGATGC 123772
RESULT 6
LOCUS      BD127258      1979 bp      DNA      linear      PAT 18-SEP-2002
DEFINITION      Primer for synthesizing full-length cDNA and use thereof.
ACCESSION      BD127258
VERSION      BD127258.1 GI:23222203
KEYWORDS      JP 2002017375-A/2689.
SOURCE      Homo sapiens (human)
ORGANISM      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 1979)
AUTHORS      Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.

```

TITLE Primer for synthesizing full-length cDNA and use thereof
 JOURNAL Patent: JP 2002017375-A 2689 22-JAN-2002;
 COMMENT HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/2689
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
 PI ISHII,
 PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TETSUJI OTSUKI, HISASHI KOGA
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/PC
 10, C12P21/02, C12Q1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 Location/Qualifiers
 FT CDS Location/Qualifiers
 1. 1979
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 451 a 542 c 474 g 512 t
 ORIGIN
 Query Match 75.7%; Score 1547; DB 6; Length 1979;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 647 CCCCTCCCTCAACCTGATATGCTGACCTGCTCCGGAGCCCCGGGCGCTGTCCGTC 706
 QY 766 CCGAGAACGCACAAAGAGAGATCTCATGATTCAGATCCGATTTGTATGGGAGCATGA 825
 DB 707 CCGAGAACGCACAAAGAGAGATCTCATGATTCAGATTCGATTTGTATGGGAGCATGA 766
 QY 826 GCAAAAACCTCAAG 885
 DB 767 GCAAAAACCTCAAG 826
 QY 886 GCTGGAGATCTCAAG 945
 DB 827 GCTGGAGATCTCAAG 886
 QY 946 TGTGCGCTATGAG 1005
 DB 887 TGTGCGCTATGAG 946
 QY 1006 CGTGGGATTTGGAATTTTGGCCCATCTTCAGAGAGAGAGAGAGAGAGAGAGAGAG 1065
 DB 947 CGTGGGATTTGGAATTTTGGCCCATCTTCAGAGAGAGAGAGAGAGAGAGAGAGAG 1006
 QY 1066 GGGCATGGGAG 1125
 DB 1007 GGGCATGGGAG 1066
 QY 1126 TTGGCGCTGCTTTTGGCCCATCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1185
 DB 1067 TTGGCGCTGCTTTTGGCCCATCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1126
 QY 1186 CATGAATTTTGTGGGCTACCGGACAGTCAAGATTTGAACAAGAGAGAGAGAGAGAG 1245
 DB 1127 CATGAATTTTGTGGGCTACCGGACAGTCAAGATTTGAACAAGAGAGAGAGAGAG 1186
 QY 1246 GGAATTTTGTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1305
 DB 1187 GGAATTTTGTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1246
 QY 1306 TGTGTCACCTGAG 1365
 DB 1247 TGTGTCACCTGAG 1306
 QY 1366 TCTCTAGAGCTTACATCAATGTTGTGGGATACACTGAGAGAGAGAGAGAGAGAG 1425
 DB 1307 TCTCTAGAGCTTACATCAATGTTGTGGGATACACTGAGAGAGAGAGAGAGAGAG 1366
 QY 1426 TGTCTAAG 1485
 DB 1367 TGTCTAAG 1426
 QY 1486 TATGTAG 1545
 DB 1427 TATGTAG 1486
 QY 1546 TCTGTGTGGGAG 1605
 DB 1487 TCTGTGTGGGAG 1546
 QY 1606 ATCAGTGAATGATCATTAACCTCCGCTGACATCTTGCCCAATGGGAGATGAT 1665
 DB 1547 ATCAGTGAATGATCATTAACCTCCGCTGACATCTTGCCCAATGGGAGATGAT 1606
 QY 1666 TTCACCAAG 1703
 DB 1607 TTCACCAAG 1664

RESULT 7
 AK074746 1979 bp mRNA linear PRI 03-SEP-2002
 LOCUS Homo sapiens cDNA FLJ90265 fis. clone NTBRM001325, weakly similar
 DEFINITION TO CHONDROITIN 6-SULFOTRANSFERASE (EC 2.8.2.17).
 ACCESSION AK074746

VERSION AK074746.1 GI:22760388
 KEYWORDS oligo capping; fls (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Mammalia; Euthera; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Iisogai,T., Ota,T., Nishikawa,T., Hayashi,K., Otsuki,T.,
 Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S.,
 Kawai,H.O.Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y.,
 Kojima,S., Nagahara,K., Masuno,Y., Ono,T., Okano,K., Yoshikawa,Y.,
 Aotsuka,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and
 Ninomiya,K.
 TITLE NEDO human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1979)
 AUTHORS Iisogai,T. and Otsuki,T.
 TITLE Direct Submission
 JOURNAL Submitted (25-MAR-2002) Takao Iisogai, Helix Research Institute,
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
 COMMENT NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan: cDNA full insert sequencing;
 Research Association for Biotechnology: cDNA library construction;
 Institute of Medical Science, University of Tokyo, Laboratory of
 Genome Structure, Human Genome Center: cDNA 5'- & 3'-end one pass
 sequencing and clone selection: Helix Research Institute (supported
 by Japan Key Technology Center etc.).
 FEATURES
 SOURCE Location/Qualifiers
 1..1979
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="NT2RM4001325"
 /cell_line="NT2"
 /cell_type="teratocarcinoma"
 /clone_id="NT2RM4"
 /note="cloning vector: PME18SFL3-mRNA from uninduced NT2
 neuronal precursor cells"
 69..1229
 /note="unamed protein product"
 /codon_start=1
 /protein_id="BAC11177.1"
 /db_xref="GI:22760389"
 /translation="MLPFRKPKDKLFLVSOMAILALFPHYSHNISLSMKAOERH
 VIASSRSGSSFFGLFGHDPFVFLMFMHMYMFKQSTAMLMAMAYRDLIRAF
 LCDMSYDAYHEBEPKROSSLFQWENRALSAPACITIPDEITIPAKRLDSQP
 FEYERKCRNISTSHVLEKVEFENQSLYPLKDSLHLVHLVDRRAVERSRERK
 GDLMDIRIYGOHEOKLEKEDQYVMOVYCOSQLYVTLQSLPRALDERYLVRY
 EDLARAVACTSRMYEFGLEFLPHLOTVHNITRGGMGDHAFHTVARALNVSAM
 RMSLPYKVKYRLKACGDANMLLGRVHRSBQBNLLDLSTWTVPEQIH"
 CDS
 ORIGIN
 BASE COUNT 451 a 542 c 474 g 512 t
 Query Match 75.7%; Score 1547; DB 9; Length 1979;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 346 GCCCGCTGACAGCTGTGATGACCTTCAAGCAGACGCCCTGATGCTGCATGCGC 405
 Db 287 GCCCGCTGACAGCTGTGATGACCTTCAAGCAGACGCCCTGATGCTGCATGCGC 346
 QY 406 TGTGGGATCTGATACGGCCGCTCTTGTGGACATGAGCTCTTTGATGCTCAT 465
 Db 347 TGTGGGATCTGATACGGCCGCTCTTGTGGACATGAGCTCTTTGATGCTCAT 406
 QY 466 GGAACCTGTGTCGCCGAGACAGTCCAGCCTCTTCACTGGGAGAACAGCGGCGCTGTG 525
 Db 407 GGAACCTGTGTCGCCGAGACAGTCCAGCCTCTTCACTGGGAGAACAGCGGCGCTGTG 466
 QY 526 TTTGCAACCTGCGCTGATATCGTGCACCTGGTCCGGAGCCCGGGGCTCACTGAG 585
 Db 467 TTTGCAACCTGCGCTGATATCGTGCACCTGGTCCGGAGCCCGGGGCTCACTGAG 526
 QY 586 GCTCTGTGACATGACAGCCCTTGTGAGTGTGAGAGAGCGCTGCTCTACAGCA 645
 Db 527 GCTCTGTGACATGACAGCCCTTGTGAGTGTGAGAGAGCGCTGCTCTACAGCA 586
 QY 646 CGTGTGCTCAAGAGGTGGCTTCTTCAACCTGCACTCCCTTACCCGCTGGAAGA 705
 Db 587 CGTGTGCTCAAGAGGTGGCTTCTTCAACCTGCACTCCCTTACCCGCTGGAAGA 646
 QY 706 CCCCTGCTCAACCTGATATCGTGCACCTGGTCCGGAGCCCGGGGCTCACTGAG 765
 Db 647 CCCCTGCTCAACCTGATATCGTGCACCTGGTCCGGAGCCCGGGGCTCACTGAG 706
 QY 766 CCGAAGACGCAAGAGGAGATCTCATGATGATGACAGTGCATGATGAGGCGAGCATGA 825
 Db 707 CCGAAGACGCAAGAGGAGATCTCATGATGATGACAGTGCATGATGAGGCGAGCATGA 766
 QY 826 GCAAAACTCAAGAGAGAGACCAACCTACTATGATGAGGTCACTTGCACAAAGCA 885
 Db 767 GCAAAACTCAAGAGAGAGACCAACCTACTATGATGAGGTCACTTGCACAAAGCA 826
 QY 886 GCTGAGATCTACAAAGACATCCAGTCTTCCCAAGGCCCTGAGAGAACGCTACTGCT 945
 Db 827 GCTGAGATCTACAAAGACATCCAGTCTTCCCAAGGCCCTGAGAGAACGCTACTGCT 886
 QY 946 TGTGCGTATGAGAGACCTGCTGAGCCCTGTGGCCAGACTTCCGAATGATGAT 1005
 Db 887 TGTGCGTATGAGAGACCTGCTGAGCCCTGTGGCCAGACTTCCGAATGATGAT 946
 QY 1006 CGTGGATTTGGAATCTTGGCCCATCTTCAAGACCTGGGTGATTAACATCCAGGCAA 1065
 Db 947 CGTGGATTTGGAATCTTGGCCCATCTTCAAGACCTGGGTGATTAACATCCAGGCAA 1006
 QY 1066 GGGCATGGGACACAGCCTTCCACACAAATGCCAGGAGTCCCTTAATGTCCTCCAGGC 1125
 Db 1007 GGGCATGGGACACAGCCTTCCACACAAATGCCAGGAGTCCCTTAATGTCCTCCAGGC 1066
 QY 1126 TTTGGCGTGTCTTTGGCCATATGAAAGTTTCTGACTTCAGAAAGCTGTGGCGATG 1185
 Db 1067 TTTGGCGTGTCTTTGGCCATATGAAAGTTTCTGACTTCAGAAAGCTGTGGCGATG 1126
 QY 1186 CATGAATTTGCTGGGCTACCGCCACGTCAGATCTGGAACAAAGAGAAACCTGTGCT 1245
 Db 1127 CATGAATTTGCTGGGCTACCGCCACGTCAGATCTGGAACAAAGAGAAACCTGTGCT 1186
 QY 1246 GGATCTTCTGCTACCTGAGATGTCCTGAGACAAATCCATTAAGAGGTTGAGAAGCTT 1305
 Db 1187 GGATCTTCTGCTACCTGAGATGTCCTGAGACAAATCCATTAAGAGGTTGAGAAGCTT 1246
 QY 1306 TGTGCGACCTGTGTGACGCTCAGTCACTTCTCTGAATGCTTCTGAGCCTTGCCTACA 1365
 Db 1247 TGTGCGACCTGTGTGACGCTCAGTCACTTCTCTGAATGCTTCTGAGCCTTGCCTACA 1306
 QY 1366 TCTCTGAGCCTTAACATACATGCTGTGTGGGTATCACTAGTGTGATTTGTGTCCACAG 1425
 Db 1307 TCTCTGAGCCTTAACATACATGCTGTGTGGGTATCACTAGTGTGATTTGTGTCCACAG 1366

1426 TGTCTAAGCAGAAAGACTTTTGTCTCATGCTTGTCTTAGAAAAACAGACTGGGGAACT 1485
1367 TGTCTAAGCAGAAAGACTTTTGTCTCATGCTTGTCTTAGAAAAACAGACTGGGGAACT 1426
1486 TATGTAGAGCAGCATCCACCAGTGAAGAGGGATTTGTCTTCTTCTTTCTTATCT 1545
1427 TATGTAGAGCAGCATCCACCAGTGAAGAGGGATTTGTCTTCTTCTTTCTTATCT 1486
1546 TCTGTCTGTGGGAGACTTTCAGAGACTTTGTGGCTGAGAGCCCTATTAAACAGACAGT 1605
1487 TCTGTCTGTGGGAGACTTTCAGAGACTTTGTGGCTGAGAGCCCTATTAAACAGACAGT 1546
1606 ATCAGTGAATGATTCATTAACCTCTCTGTCTACATCTTTGCCAATGGGGAATGATCT 1665
1547 ATCAGTGAATGATTCATTAACCTCTCTGTCTACATCTTTGCCAATGGGGAATGATCT 1606
1666 TTCACCAAGAGCTCACAGACTTTTCCACAGAGATGC 1703
1607 TTCACCAAGAGCTCACAGACTTTTCCACAGAGATGC 1644

RESULT 8
AR203335
LOCUS AR203335 2032 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 2 from patent US 6365365.
ACCESSION AR203335
VERSION AR203335.1 GI:21499698
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2032)
AUTHORS Bistup,A., Rosen,S.D., Tangemann,K. and Hemmerich,S.
TITLES Method of determining whether an agent modulates glycosyl
sulfotransferase-3
JOURNAL Patent: US 6365365-A 2 02-APR-2002;
FEATURES
source Location/Qualifiers
BASE COUNT 468 a 569 c 490 g 505 t
ORIGIN

Query Match 75.7%; Score 1547; DB 6; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 106 CAAGGCTTTCAGACTTCAGACAATGCTACTGCTTAATAAATGAAGCTCTGCTTTCT 165
160 CAAGGCTTTCAGACTTCAGACAATGCTACTGCTTAATAAATGAAGCTCTGCTTTCT 219
QY 166 GGTTCCTCCAGATGGCATCTTGGCTTATTTCTTCCACATGTACAGCCACACATCAGCTC 225
Db 220 GGTTCCTCCAGATGGCATCTTGGCTTATTTCTTCCACATGTACAGCCACACATCAGCTC 279
QY 226 CCTGTCTATGAAGGACAGCCGAGGAGCATGACGCTGCTTCTTCTGGGGCTC 285
Db 280 CCTGTCTATGAAGGACAGCCGAGGAGCATGACGCTGCTTCTTCTGGGGCTC 339
QY 286 TGGCTCTTCTTGTGTGGGAGCTTTTGGGAGCACCAGATGTTTCTTACCTGATGA 345
Db 340 TGGCTCTTCTTGTGTGGGAGCTTTTGGGAGCACCAGATGTTTCTTACCTGATGA 399
QY 346 GCGCCGCTGGCAGCTGTGATGACTTCAAGCAGACACCGCTGATCTGCACATGCG 405
Db 400 GCGCCGCTGGCAGCTGTGATGACTTCAAGCAGACACCGCTGATCTGCACATGCG 459
QY 406 TGTGGGAGATCTGATACGGGCGCTTCTTGTGGCAGATGAGGCTTGTGATGCTCAAT 465
Db 460 TGTGGGAGATCTGATACGGGCGCTTCTTGTGGCAGATGAGGCTTGTGATGCTCAAT 519
QY 466 GGAAGCTGTCTCCCGAGACAGTCCAGCTTTTTCAGTGGGAGAACAGCCGGGCTGTG 525
Db 520 GGAAGCTGTCTCCCGAGACAGTCCAGCTTTTTCAGTGGGAGAACAGCCGGGCTGTG 579

QY 526 TTCTGACCTGCCGTGACATCATCCACAAAGATGAATCATCCCGGGGCTCACTGAG 585
Db 580 TTCTGACCTGCCGTGACATCATCCACAAAGATGAATCATCCCGGGGCTCACTGAG 639
QY 586 GCTCCGTGTACAGTCAACACACCCCTTTAGAGTGTGTGAGAAAGGCTGCCGCTCCACAGCA 645
Db 640 GCTCCGTGTACAGTCAACACACCCCTTTAGAGTGTGTGAGAAAGGCTGCCGCTCCACAGCA 699
QY 646 CGTGTGTCTCAAGAGAGTGCCTTTCTTCAACCTGCAGTCCCTCTACCCGCTGTGAAGA 705
Db 700 CGTGTGTCTCAAGAGAGTGCCTTTCTTCAACCTGCAGTCCCTCTACCCGCTGTGAAGA 759
QY 706 CCCCTCCCTCAACCTGCATATGCTGACCTGTGCGGAGCCCGGGGCGGTTCGTTG 765
Db 760 CCCCTCCCTCAACCTGCATATGCTGACCTGTGCGGAGCCCGGGGCGGTTCGTTG 819
QY 766 CCGAGAGACGACAAAGAGAGATCTCATGATGTACATCCGATTTGTATGGGAGACATGA 825
Db 820 CCGAGAGACGACAAAGAGAGATCTCATGATGTACATCCGATTTGTATGGGAGACATGA 879
QY 826 GCAAAAACCTCAAGAAAGAGAGACCAACCTACTATGTGATGACAGTATGCTATGCCAAAGCA 885
Db 880 GCAAAAACCTCAAGAAAGAGAGAGACCAACCTACTATGTGATGACAGTATGCTATGCCAAAGCA 939
QY 886 GCTGAGATCTACAAAGACCATCCAGTCTTGGCCCAAGGCCCTGCAGAAAGCTTACCTGCT 945
Db 940 GCTGAGATCTACAAAGACCATCCAGTCTTGGCCCAAGGCCCTGCAGAAAGCTTACCTGCT 999
QY 946 TGTGCTGTATGAGAGACTGCTGCTGACCCCTGTGGGCCAGACTTCCCGAATGTATGAAT 1005
Db 1000 TGTGCTGTATGAGAGACTGCTGCTGACCCCTGTGGGCCAGACTTCCCGAATGTATGAAT 1059
QY 1006 CGTGGATTTGAATTTCTGCCCATCTGTGACAGCTGGTGACATTAACATCACCAGAGCA 1065
Db 1060 CGTGGATTTGAATTTCTGCCCATCTGTGACAGCTGGTGACATTAACATCACCAGAGCA 1119
QY 1066 GGGCATGGGTGACACACGCTTTCCACACAAATGCGAGATGCCCTTAATGTCTCCAGGC 1125
Db 1120 GGGCATGGGTGACACACGCTTTCCACACAAATGCGAGATGCCCTTAATGTCTCCAGGC 1179
QY 1126 TTGGGCTGCTGCTTTCCCTATGAAAGGTTTCTGACTTCAGAAAGCCGTGGGAGTGC 1185
Db 1180 TTGGGCTGCTGCTTTCCCTATGAAAGGTTTCTGACTTCAGAAAGCCGTGGGAGTGC 1239
QY 1186 CATGAATTTGTGGGCTACCGGACGTCAGATCTGAACAAAGACAGAAAGCTGTGTCT 1245
Db 1240 CATGAATTTGTGGGCTACCGGACGTCAGATCTGAACAAAGACAGAAAGCTGTGTCT 1299
QY 1246 GGAATCTTGTCTACTGACATGTCCCTGACCAATCCACTAAGAGGTTGAGAGGCTT 1305
Db 1300 GGAATCTTGTCTACTGACATGTCCCTGACCAATCCACTAAGAGGTTGAGAGGCTT 1359
QY 1306 TGTGTCACCTGCTGTCACGCTTCAGTCACTTTCTGATGAGTCTTGAGCCTTGCTTACA 1365
Db 1360 TGTGTCACCTGCTGTCACGCTTCAGTCACTTTCTGATGAGTCTTGAGCCTTGCTTACA 1419
QY 1366 TCTCTGAGCTTAATCATATGCTGTGGGTATCACACTGATGTGATGTGTCCACAG 1425
Db 1420 TCTCTGAGCTTAATCATATGCTGTGGGTATCACACTGATGTGATGTGTGTCCACAG 1479
QY 1426 TGTCTAAGCAGAAAGACTTTTGTGTCAATGCTTTGTGTCTAGAAAAACAGCTGGGAACCT 1485
Db 1480 TGTCTAAGCAGAAAGACTTTTGTGTCAATGCTTTGTGTCTAGAAAAACAGCTGGGAACCT 1539
QY 1486 TATGTAGAGCAGCATCCACCAGTGAAGAGGGATTTGTCTTCTTCTTTCTTATCT 1545
Db 1540 TATGTAGAGCAGCATCCACCAGTGAAGAGGGATTTGTCTTCTTCTTTCTTATCT 1599
QY 1546 TCTGTCTGTGGGAGACTTTCAGAGACTTTGTGGCTGAGAGCCCTATTAAACAGACAGACT 1605
Db 1600 TCTGTCTGTGGGAGACTTTCAGAGACTTTGTGGCTGAGAGCCCTATTAAACAGACAGACT 1659


```

QY 1606 ATCATGGAATTGATCCATAAACCTCCCTGTCACATCTGCCCAATGGGAATGATCT 1665
DB 1660 ATCGTGAATTGATCCATAAACCTCCCTGTCACATCTGCCCAATGGGAATGATCT 1719
QY 1666 TTCACCAAGAGCTCACCAGCATTTCACAGAGATGC 1703
DB 1720 TTCACCAAGAGCTCACCAGCATTTCACAGAGATGC 1757

RESULT 9
BD134772
LOCUS BD134772 2032 bp DNA linear PAT 18-SEP-2002
DEFINITION Glycosylsulfotransferase-3.
ACCESSION BD134772
VERSION BD134772.1 GI:23229717
KEYWORDS JP 2002507409-A/1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Bistrup, A., Rosen, S.D., Tangemann, K. and Hemmerich, S.
Glycosylsulfotransferase-3
Patent: JP 2002507409-A 1 12-MAR-2002;
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA, SYNTAX INC
OS Homo sapiens (human)
PN JP 2002507409-A/1
PD 12-MAR-2002
PF 26-FEB-1999 JP 2000537979
PR 20-MAR-1998 US 09/045284,12-NOV-1998 US 09/190911 PT
ANNEETTE BISTRUP, STEVEN D ROSEN, KIRSTEN TANGEMANN, STEFAN PI
HEMMERICH
PC C12N9/10,A01K67/027,A61K45/00,A61K48/00,A61P29/00,A61P37/06,
PC A61P43/00,
PC C12N1/15,C12N1/21,C12N5/10,C12N15/09,C12Q1/48,C12N5/00,C12N15/
PC 00
CC Glycosylsulfotransferase-3
CCH key Location/Qualifiers
FH source 1..2032
FT Location/Qualifiers
1..2032
/organism="Homo sapiens (human)".

FEATURES
Source
1..2032
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT 468 a 569 c 490 g 505 t

Query Match 75.7%; Score 1547; DB 6; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

DB 460 TGTGCGGATCTGATGATGAGGCGCTCTTGTGACATGAGCTCTTGTGATGCTACAT 519
QY 466 GGAACCTGTGTCCCGGAGACAGTCCAGCTCTTCACTGGGAGAACAGCGGGCCCTGTG 525
DB 520 GGAACCTGTGTCCCGGAGAGACAGTCCAGCTCTTCACTGGGAGAACAGCGGGCCCTGTG 579
QY 526 TTTGTGACAGCTGCGGTGATCATCATCCACAAGATGAATATCCCGGGGCTCACTGAG 585
DB 580 TTTGTGACAGCTGCGGTGATCATCATCCACAAGATGAATATCCCGGGGCTCACTGAG 639
QY 586 GCTCTGTGACATGACAGCCCTTTGAGTGTGGAGAGAGCCCTGCGCTCTACAGCA 645
DB 640 GCTCTGTGACATGACAGCCCTTTGAGTGTGGAGAGAGCCCTGCGCTCTACAGCA 699
QY 646 GCTGTGTCTCAAGAGAGTGGCTCTTCAACCTCACTGCTCTCACTCCGCTGTGAAGA 705
DB 700 GCTGTGTCTCAAGAGAGTGGCTCTTCAACCTCACTGCTCTCACTCCGCTGTGAAGA 759
QY 706 CCCCTCCCTCAACCTGATATGTCACCTGCTCGGGAGACCCCGGGCCGTTCGGTTC 765
DB 760 CCCCTCCCTCAACCTGATATGTCACCTGCTCGGGAGACCCCGGGCCGTTCGGTTC 819
QY 766 CCGAGAACGACAAAGGAGATCTCATGATTGACAGTCCGATTGTGATGGGACAGATGA 825
DB 820 CCGAGAACGACAAAGGAGATCTCATGATTGACAGTCCGATTGTGATGGGACAGATGA 879
QY 826 GCAAAAACCTCAAGAAAGAGAGCCCAACCTACTATGATGATGAGTGCATCTGCCAAAGCA 885
DB 880 GCAAAAACCTCAAGAAAGAGAGCCCAACCTACTATGATGATGAGTGCATCTGCCAAAGCA 939
QY 886 GCTGGAATCTCAAGAGACATCATAGCTCTTCCCAAGGAGCCCTGAGAGACCTGCTCT 945
DB 940 GCTGGAATCTCAAGAGACATCATAGCTCTTCCCAAGGAGCCCTGAGAGACCTGCTCT 999
QY 946 TGTGCGCTGATGAGAGACCTGCTGAGGCGCTGTGGCCAGACTTCCGAATGTATGAT 1005
DB 1000 TGTGCGCTGATGAGAGACCTGCTGAGGCGCTGTGGCCAGACTTCCGAATGTATGAT 1059
QY 1006 CGTGGGATGGAATCTTGTGCCCCCATCTTCAACACCTGGGTGATTAATCAACCCAGCA 1065
DB 1060 CGTGGGATGGAATCTTGTGCCCCCATCTTCAACACCTGGGTGATTAATCAACCCAGCA 1119
QY 1066 GGGCATGGGTACACAGCTTCCACACAAATGCCAGGAGTCCCTTAATGTCTCCAGGC 1125
DB 1120 GGGCATGGGTACACAGCTTCCACACAAATGCCAGGAGTCCCTTAATGTCTCCAGGC 1179
QY 1126 TTTGCGCTGTGCTTTTGGCCCTATGAAAGGTTTCTGACTTCAAGAAAGCTGTGCGATGC 1185
DB 1180 TTTGCGCTGTGCTTTTGGCCCTATGAAAGGTTTCTGACTTCAAGAAAGCTGTGCGATGC 1239
QY 1186 CATGAATTTGCTGGGCTACCGCCACAGTCAGATCTGGAACAGAAAGAAACCTGTGCT 1245
DB 1240 CATGAATTTGCTGGGCTACCGCCACAGTCAGATCTGGAACAGAAAGAAACCTGTGCT 1299
QY 1246 GGATCTTCTGTCTACCTGAGATGTCCTGAGCAATACACATAGAGGTTGAGAAGCTT 1305
DB 1300 GGATCTTCTGTCTACCTGAGATGTCCTGAGCAATACACATAGAGGTTGAGAAGCTT 1359
QY 1306 TGTGCGACAGCTGTGTCAGCTCAGTCACTTCTCTGATGCTCTGAGCTTCCCTTACA 1365
DB 1360 TGTGCGACAGCTGTGTCAGCTCAGTCACTTCTCTGATGCTCTGAGCTTCCCTTACA 1419
QY 1366 TCTCTGAGCCTTAACTACATGTCGTGGGATATCACTGAGTGTGATGCTGTCACAGC 1425
DB 1420 TCTCTGAGCCTTAACTACATGTCGTGGGATATCACTGAGTGTGATGCTGTCACAGC 1479
QY 1426 TGTCTCAAGCAAGAGACTTGTGTGCTCATGCTTGTCTGAAAGAGAGCTGGGAGACT 1485
DB 1480 TGTCTCAAGCAAGAGACTTGTGTGCTCATGCTTGTCTGAAAGAGAGCTGGGAGACT 1539
QY 1486 TATGTGAGAGACATCTCCACAGTGAAGAGGTTATGTGCTCTTCTTGTGATCT 1545

```

Db	1540	TATGAGACACACATCCACCAGTGAANAGGGATGTGCCTTCCTCCTTTCTTGACT	1599
Qy	1546	TCCTGTCTGGGACACTTCACAGACTTTGTGGCCCTGAGGCCATTAAAGCACACAGT	1605
Db	1600	TCCTGTCTGGGACAGCTTCACAGACTTTGTGGCCCTGAGGCCATTAAAGCACACAGT	1659
Qy	1606	ATCATGTGAATTGATCAATTAACGCCCCCTGTCCACATATTCGCCAATGGGAAATGATCT	1665
Db	1660	ATCATGTGAATTGATCAATTAACGCCCCCTGTCCACATATTCGCCAATGGGAAATGATCT	1719
Qy	1666	TTACACCAAAGAGCTCACCGACATTTTCCACAGAGATGC	1703
Db	1720	TTACACCAAAGAGCTCACCGACATTTTCCACAGAGATGC	1757
RESULT 10			
AF131235			
DEFINITION	Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA,	2032 bp	PRI 10-AUG-1999
ACCESSION	complete cds.		
VERSION	AF131235		
KEYWORDS	AF131235.1 GI:4927113		
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 2032) Bistrup,A., Bhakta,S., Lee,J.K., Belov,Y.Y., Gunn,M.D., Zuo,F.R., Huang,C.C., Kannagi,R., Rosen,S.D. and Hemmerich,S. Sulfotransferases of two specificities function in the reconstitution of high endothelial cell ligands for L-selectin J. Cell Biol. 145 (4), 899-910 (1999)		
TITLE	JOURNAL MEDLINE		
REFERENCE	PUBMED		
AUTHORS	2 (bases 1 to 2032) Bistrup,A., Tangemann,K., Bhakta,S., Lee,J.-K., Belov,Y.Y., Gunn,M.D., Zuo,F.-R., Huang,C.-C., Kannagi,R., Rosen,S.D. and Hemmerich,S. Direct Submission Submitted (25-FEB-1999) Respiratory Diseases, Roche Bioscience, 3401 Hillview Avenue, Palo Alto, CA 94304, USA Location/Qualifiers		
FEATURES	source		
CDS	1..2032 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /cell_type="high endothelial" /tissue_type="tonsil" 182..1342 /function="biosynthesis of endothelial ligands for L-selectin" /note="HEC-GlCNAc-6-ST" /codon_start=1 /product="N-acetylglucosamine 6-O-sulfotransferase" /protein_id="AAD33015.1" /db_xref="GI:4927114" translation="MLPCKKKKLFLVSOAILALFFHWYSHNISSLMKAQERHH LVIVLSMSGGSFVGGLFGHPDYELMEPRAMWMMFKOSTAMILMAAYDLIRAF LCDSVFDAHYEPPRGOSLFQENSRSLASAPCIDITODEITIPAHCLSDQP FEVEAKCRSTSHVYLKEVRFNLOSITPLIKDSLNLDVHLVDRAVFRSERKRR GDMLIDRSIVWGQHOKLKIEDQYYVMVTCOSOLIKTKTGISLPALDERYLAVY EDLARPAQFSRMTEFYEGLETFPHLOTVINIITRGGMGHAFHTNARALNVSAWM RMSIPYEVSRLQACGDAMNLIGRYHRVSRBOQRNLLDLSITWTVPEDQH"		
BASE COUNT	468 a 569 c 490 g 505 t		
ORIGIN			
Query Match	75.7%; Score 1547;	DB 9;	Length 2032;
Best Local Similarity	99.9%;	Pred. No. 0;	
Matches 1597;	Conservative 0;	Mismatches 1;	Indels 0; Gaps 0;
QY	106	CAAGTGTTCACATTCACACACAATGCTACTGCTAAAAAATGAAGCTCCTGCTGTTCT	165

Db	160	CAGGTCCTTCACATTCACACACAATGCTACTGCTCTAAAAAAATGAAAGCTTCGCTCTTTCT	219
QY	166	GGTTTCCAGAGATGGCCATCTTGAGCTATTTCTTCACATGTACAGCCCAACATCAGCTC	225
Db	220	GGTTTCCAGATGGCCATCTTGAGCTATTTCTTCACATGTACAGCCCAACATCAGCTC	279
QY	226	CCTGCTATGAAAGGACAGCCCGAGGCGATCAGCTGCTGTTGCTTCTCTGGCGCTC	285
Db	280	CCTGCTATGAAAGGACAGCCCGAGGCGATCAGCTGCTGTTGCTTCTCTGGCGCTC	339
QY	286	TGGCTCTTCTTTTGTGGGGGAGCTTTTGTGGGAGCAGCCAGATGTTTTCTACTGATGGA	345
Db	340	TGGCTCTTCTTTTGTGGGGGAGCTTTTGTGGGAGCAGCCAGATGTTTTCTACTGATGGA	399
QY	346	GCCCGCCCTGGGACAGCTGTGGATGACCTTTAAACAGAGACCCGCTGTGATGCTGCACATGGC	405
Db	400	GCCCGCCCTGGGACAGCTGTGGATGACCTTTAAACAGAGACCCGCTGTGATGCTGCACATGGC	459
QY	406	TGTGGGAGTATGATATACGGGCGGCTTTCTTGTGCACATGACGCTTTGATAGGCTACAT	465
Db	460	TGTGGGAGTATGATATAGGGGCGGCTTTCTTGTGCACATGAGCGCTTTGATAGGCTACAT	519
QY	466	GGAACCTGTGTCCTCCGGAGACAGTCCAGCCTCTTTACGTGGAGAAACACCGGGCCCTGTG	525
Db	520	GGAACCTGTGTCCTCCGGAGACAGTCCAGCCTCTTTACGTGGAGAAACACCGGGCCCTGTG	579
QY	526	TTTCTGCACCTCCCTGTGACATCATCCCAAGATATGAATTCATCCCGCGGTCACTGAG	585
Db	580	TTTCTGCACCTCCCTGTGACATCATCCCAAGATATGAATTCATCCCGGGGTCACTGAG	639
QY	586	GCTCTCTGTGCAGTCAACAGCCCTTTGAGGTGTGGAGAAAGCCCTGCGCTCTTACAGCCA	645
Db	640	GCTCTCTGTGCAGTCAACAGCCCTTTGAGGTGTGGAGAAAGCCCTGCGCTCTTACAGCCA	699
QY	646	CGTGGTGTCAAGGAGTGCGCTTCTTCAACTGTGAGTCCCTACCCGCTGTCTGAAAGA	705
Db	700	CGTGGTGTCAAGGAGTGCGCTTCTTCAACTGTGAGTCCCTACCCGCTGTCTGAAAGA	759
QY	706	CCGCTCCTCAACCTGCATATGTGTGCACCTGTGTCGGGAGCCCGGGCGTGTTCGTTG	765
Db	760	CCGCTCCTCAACCTGCATATGTGTGCACCTGTGTCGGGAGCCCGGGCGTGTTCGTTG	819
QY	766	CCGAGAACGCAAAAGGAGATATCATATTGACAGTGTGCAGTGTGGAGGGGACAGATGA	825
Db	820	CCGAGAACGCAAAAGGAGATATCATATTGACAGTGTGCAGTGTGGAGGGGACAGATGA	879
QY	826	GCAAAAACCTCAAGAAAGAGGACCAACCTTACTATGTGTGACAGGTCACTGCCAAAGCCA	885
Db	880	GCAAAAACCTCAAGAAAGAGGACCAACCTTACTATGTGTGACAGGTCACTGCCAAAGCCA	939
QY	886	GCTGGAGATTCACAAAGACATCCAGTCTCTCCCAAGGCGCTGCAGAGAACGCTACGCT	945
Db	940	GCTGGAGATTCACAAAGACATCCAGTCTCTCCCAAGGCGCTGCAGAGAACGCTACGCT	999
QY	946	TGTGGCTATGAGGACGCTGGCTGCAGGCCCTGTGGCCGACAGCTTCCCAATATGTAAT	1005
Db	1000	TGTGGCTATGAGGACGCTGGCTGCAGGCCCTGTGGCCGACAGCTTCCCAATATGTAAT	1059
QY	1006	CGTGGAGATTCGAATTTCTGGCCCATCTTCAAGACCTGGGTGCATACATCACCCAGAGCAA	1065
Db	1060	CGTGGAGATTCGAATTTCTGGCCCATCTTCAAGACCTGGGTGCATACATCACCCAGAGCAA	1119
QY	1066	GGGCACTGGGTACCAACGCTTTCACACAAATGSCAGGGATCCCTTAATGTCTCCAGGC	1125
Db	1120	GGGCACTGGGTACCAACGCTTTCACACAAATGSCAGGGATCCCTTAATGTCTCCAGGC	1179
QY	1126	TTGGGCGTGTCTTTGGCCCTATGAAAAGTTTCTGCATTCAGAAAAGCCTGTGGCGATGC	1185
Db	1180	TTGGGCGTGTCTTTGGCCCTATGAAAAGTTTCTGCATTCAGAAAAGCCTGTGGCGATGC	1239
QY	1186	CATGAATTTGCTGGGCTACCGCCACAGTCAAGTCTGAAACAAACAGAAACCTGTGTCT	1245
Db	1240	CATGAATTTGCTGGGCTACCGCCACAGTCAAGTCTGAAACAAACAGAAACCTGTGTCT	1299

QY 1246 GGATCTTCTGTCTACTGAGTGTGCTCCGAGCAATCCACTAAGAGGTTGAGAAGCCTT 1305
 DB 1300 GGATCTTCTGTCTACTGAGTGTGCTCCGAGCAATCCACTAAGAGGTTGAGAAGCCTT 1359
 QY 1306 TGTGCGACCTGTGTGTCAGCTTCTGTGTAAGTCTTGTGAGCCTTGTGCTTACA 1365
 DB 1360 TGTGCGACCTGTGTGTCAGCTTCTGTGTAAGTCTTGTGAGCCTTGTGCTTACA 1419
 QY 1366 TGTGCGACCTGTGTGTCAGCTTGTGTAAGTCTTGTGAGCCTTGTGCTTACA 1425
 DB 1420 TGTGCGACCTGTGTGTCAGCTTGTGTAAGTCTTGTGAGCCTTGTGCTTACA 1479
 QY 1426 TGTGCGACCTGTGTGTCAGCTTGTGTAAGTCTTGTGAGCCTTGTGCTTACA 1485
 DB 1480 TGTGCGACCTGTGTGTCAGCTTGTGTAAGTCTTGTGAGCCTTGTGCTTACA 1539
 QY 1486 TGTGCGACCTGTGTGTCAGCTTGTGTAAGTCTTGTGAGCCTTGTGCTTACA 1545
 DB 1540 TGTGCGACCTGTGTGTCAGCTTGTGTAAGTCTTGTGAGCCTTGTGCTTACA 1599
 QY 1546 TGTGCGACCTGTGTGTCAGCTTGTGTAAGTCTTGTGAGCCTTGTGCTTACA 1605
 DB 1600 TGTGCGACCTGTGTGTCAGCTTGTGTAAGTCTTGTGAGCCTTGTGCTTACA 1659
 QY 1606 TGTGCGACCTGTGTGTCAGCTTGTGTAAGTCTTGTGAGCCTTGTGCTTACA 1665
 DB 1660 TGTGCGACCTGTGTGTCAGCTTGTGTAAGTCTTGTGAGCCTTGTGCTTACA 1719
 QY 1666 TGTGCGACCTGTGTGTCAGCTTGTGTAAGTCTTGTGAGCCTTGTGCTTACA 1703
 DB 1720 TGTGCGACCTGTGTGTCAGCTTGTGTAAGTCTTGTGAGCCTTGTGCTTACA 1757

RESULT 11
 AFI49783 1333 bp mRNA linear PRI 02-JUL-2001
 LOCUS Homo sapiens L-selectin ligand sulfotransferase mRNA, complete cds.
 DEFINITION AFI49783
 ACCESSION AFI49783.1 GI:13897503
 VERSION AFI49783.1
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS Yeh, J.C., Hiraoka, N., Petryniak, B., Nakayama, J., Elies, L.G., Rabuka, D., Hindsgraul, O., Marth, J.D., Lowe, J.B. and Fukuda, M.
 TITLE Novel sulfated lymphocyte homing receptors and their control by a Core1 extension beta 1,3-N-acetylglycosaminyltransferase
 JOURNAL Cell 105 (7), 957-969 (2001)
 MEDLINE 21332592
 PUBMED 11439191
 REFERENCE 2 (bases 1 to 1333)
 AUTHORS Hiraoka, N. and Fukuda, M.
 TITLE Direct Submission
 JOURNAL Submitted (10-MAY-1999) Glycobiology, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

FEATURES
 Source
 1. 1333
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 93. 1253
 /note="LSSR"
 /codon_start=1
 /product="L-selectin ligand sulfotransferase"
 /protein_id="AAK48417.1"
 /db_xref="GI:13897504"
 /translation="MLPKKKKLLFLVSOAAILALFPHYSHNLSISMAKQPERH
 VIVLSWRSGSSFYGOLEFGOHDPVFLMEPRNHWMTFKOSTAMLMAYADILRAYF
 LQMSVDAITHEPRKRSALFLQWENSRALSDAPACDITPDELTIPRHCHLLCSQP
 FEVEKACRSYSHVLEKVEFREFNLSLYPLLKPSLMLHIVHLVDRPRAVRSRERIK

BASE COUNT 286 a 393 c 336 g 318 t
 ORIGIN
 Query Match 61.8%; Score 1262; DB 9; Length 1333;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 107 AAGCTCTTCACTTACGACACATGCTACTGCTTAAAAAATGAAGCTCTCTGTTCTG 166
 DB 72 AAGCTCTTCACTTACGACACATGCTACTGCTTAAAAAATGAAGCTCTCTGTTCTG 131
 QY 167 GTTTCACAGATGGGCTTGGCTTCTTCTTCCAGATGTCACGCAACATACAGCTCC 226
 DB 132 GTTTCACAGATGGGCTTGGCTTCTTCTTCCAGATGTCACGCAACATACAGCTCC 191
 QY 227 CTGTCTATGAAGGACAGCCCGAGCGATGACGCTGTGCTTCTGCTGCGCTCT 286
 DB 192 CTGTCTATGAAGGACAGCCCGAGCGATGACGCTGTGCTTCTGCTGCGCTCT 251
 QY 287 GGTCTCTTCTTGTGGGACCTTGTGGGACGACCCAGATGTTTCTTACCTGATGAG 346
 DB 252 GGTCTCTTCTTGTGGGACCTTGTGGGACGACCCAGATGTTTCTTACCTGATGAG 311
 QY 347 CCGGCTTGGACAGTGTGATACCTTCAAGACAGACCCGCTGATGCTCAATGCT 406
 DB 312 CCGGCTTGGACAGTGTGATACCTTCAAGACAGACCCGCTGATGCTCAATGCT 371
 QY 407 GTGGGGATGTGATACGGGCGGTCTTGTGTGCAACATGACGCTTGTGATGCTTAC 466
 DB 372 GTGGGGATGTGATACGGGCGGTCTTGTGTGCAACATGACGCTTGTGATGCTTAC 431
 QY 467 GAACCTGTGCTCCGAGACAGTCAGCCTTCTTTCAGTGGAGAAACGCGGCGCTGTGT 526
 DB 432 GAACCTGTGCTCCGAGACAGTCAGCCTTCTTTCAGTGGAGAAACGCGGCGCTGTGT 491
 QY 527 TGTGACCTGCTGTGACATCATCCCAACATGAATGATCCCGGGGCTCACTGACAG 586
 DB 492 TGTGACCTGCTGTGACATCATCCCAACATGAATGATCCCGGGGCTCACTGACAG 551
 QY 587 CTCCTGTGACATCAACAGCCCTTGTGAGTGTGAGAGAGGCGCTCCGCTCTACAGCCAC 646
 DB 552 CTCCTGTGACATCAACAGCCCTTGTGAGTGTGAGAGAGGCGCTCCGCTCTACAGCCAC 611
 QY 647 GTGCTGTCAAGAGAGTGTGCTTCAACCTGCAAGTCCCTCAACCGCTGTGAAGAC 706
 DB 612 GTGCTGTCAAGAGAGTGTGCTTCAACCTGCAAGTCCCTCAACCGCTGTGAAGAC 671
 QY 707 CCGTCCCAACCTGCAAGTGTGACCTGTGCGGGGACCCCGGGGCGGTTCGTTCC 766
 DB 672 CCGTCCCAACCTGCAAGTGTGACCTGTGCGGGGACCCCGGGGCGGTTCGTTCC 731
 QY 767 CGAAGACGACAAAGGAGATCTCATGTTTACAGTGCATGTTGATGGGACAGATGAG 826
 DB 732 CGAAGACGACAAAGGAGATCTCATGTTTACAGTGCATGTTGATGGGACAGATGAG 791
 QY 827 CAAAACCTCAAGAGAGAGCAACCTCTATGATGACAGTGCATCTGCCAAGCCAG 886
 DB 792 CAAAACCTCAAGAGAGAGCAACCTCTATGATGACAGTGCATCTGCCAAGCCAG 851
 QY 887 CTGAGAGTCTCAAGAGACATCATGCTTGTGCGCAAGGCGCTGCAAGAGAGCTTACTGCTT 946
 DB 852 CTGAGAGTCTCAAGAGACATCATGCTTGTGCGCAAGGCGCTGCAAGAGAGCTTACTGCTT 911
 QY 947 GTGGCTATGAGGACCTGTGCTGAGCCCTGTGCGCAAGATTCGCCAATGTATGAATTC 1006
 DB 912 GTGGCTATGAGGACCTGTGCTGAGCCCTGTGCGCAAGATTCGCCAATGTATGAATTC 971
 QY 1007 GTGGATTTGGAATTTCTTCCCATCTTCAAGCTGTGGTGCATCAACATCACCCGAGCAAG 1066
 DB 972 GTGGATTTGGAATTTCTTCCCATCTTCAAGCTGTGGTGCATCAACATCACCCGAGCAAG 1031

QY	1067	GGCATGGGGTGAACACGGCTTTCCACAACAATAACCAAGGGATGGCCCTTAATGTCCTCCAGGCT	1126
Db	1032	GGCATGGGGTGAACACGGCTTTCCACAACAATAACCAAGGGATGGCCCTTAATGTCCTCCAGGCT	1091
QY	1127	TGGCGCTGTCCTTTGGCCCTATGAAGAAAGGTTTCTGACATTCAGAAAGCCTGTGGCGATGCG	1186
Db	1092	TGGCGCTGTCCTTTGGCCCTATGAAGAAAGGTTTCTGACATTCAGAAAGCCTGTGGCGATGCG	1151
QY	1187	ATGAATTTTGGCTGGGCTACCGCCACGTCAGATCTGACACAAGAACAGAGAAACCTGTGCTG	1246
Db	1152	ATGAATTTTGGCTGGGCTACCGCCACGTCAGATCTGACACAAGAACAGAGAAACCTGTGCTG	1211
QY	1247	GATTTCTGTCTTACCTGGACTGTGCTCCCTGAGCAATTCACCTAAGAGGGTTGAGAAAGCTTT	1306
Db	1212	GATTTCTGTCTTACCTGGACTGTGCTCCCTGAGCAATTCACCTAAGAGGGTTGAGAAAGCTTT	1271
QY	1307	GCTGCCACCTGGTGTACAGCCTCAGTTCACCTTCTGTGAATGGTTGTGAGCCCTGGCTACAT	1366
Db	1272	GCTGCCACCTGGTGTACAGCCTCAGTTCACCTTCTGTGAATGGTTGTGAGCCCTGGCTACAT	1331
QY	1367	CT 1368	
Db	1332	CT 1333	

RESULT	12			
LOCUS	AR310485			
DEFINITION	Sequence	26 from patent US 6558935.	DNA	linear
ACCESSION	AR310485			
VERSION	AR310485.1	GI:31703448		
KEYWORDS				
SOURCE	Unknown.			

ORGANISM	Unknown.
	Unclassified.
REFERENCE	1 (bases 1 to 2065)
AUTHORS	Tang,X.T., Corley,N.C., Guejler,K.J., Baughn,M.R., Tal,P., Yue,H., Hillman,O.L. and Azilamal,Y.
JOURNAL	Human transferrase-proteins
FEATURES	Patent: US 6558935-A 26 06-MAY-2003;
source	Location/Qualifiers
	1..2065
	/organism="unknown"
BASE COUNT	466 a 575 c 491 g 533 t
ORIGIN	
Query Match	56.4%; Score 1153; DB 6; Length 2065;
Best Local Similarity	99.6%; Pred. No. 0;
Matches 1593; Conservative	0; Mismatches 4; Indels 2; Gaps 2;

[illegible]

Db	452	TTGTCGGGAACTGATTAACGGGCGCCGCTTCTTGTGGACATGAGCGCTCTTGTGACCTTACAT	511
Qy	466	GGAAACCTGGTCCCCCGGAGACAGTCCACGCTTTTCACTGGAGAAACAAGCCGGGCCCTGTG	525
Db	512	GGAACTGGTGTCCCGGAGACAGTCCACGCTTTTCACTGGAGAAACAAGCCGGGCCCTGTG	571
Qy	526	TTTCAGCACTGCTGTGATCATCATCCACAAGATGAATCATCCCCGGGGGTCACTGCAG	585
Db	572	TTTCAGCACTGCTGTGATCATCATCCACAAGATGAATCATCCCCGGGGGTCACTGCAG	630
Qy	586	GCTCCTGTGACGTAAACAGCCCTTTG-AGTGGTGGAGAAAGGCGTGGCCCTCTCTACAGCC	644
Db	631	GCTCCTGTGACGTAAACAGCCCTTTGAAAGTGGTGGAGAAAGCGTGGCCCTCTCTACAGCC	690
Qy	645	ACGTGGTGTCAAGAGAGGTGGCGCTTCTTCAACCTGCAGTCCCTCTACCCGCTGTGAAG	704
Db	691	ACGTGGTGTCAAGAGAGGTGGCGCTTCTTCAACCTGCAGTCCCTCTACCCGCTGTGAAG	750
Qy	705	ACCCCTCCCTCAACCTGCATATGCTGCACTGGTCCGGAGCCCCGGGGCGGTGTCCGTT	764
Db	751	ACCCCTCCCTCAACCTGCATATGCTGCACTGGTCCGGAGCCCCGGGGCGGTGTCCGTT	810
Qy	765	CCCGAAGACGCACAAGAGGAGATGTCATGATTTGACAGTCGGATTTGTATGGAGGAGCATG	824
Db	811	CCCGAAGACGCACAAGAGGAGATGTCATGATTTGACAGTCCGATTTGTATGGAGGAGCATG	870
Qy	825	AGCAAAAACCTCAAGAGGAGGACCAACCCCTAATCTATGTGATGACAGTCACTGCCAAAACC	884
Db	871	AGCAAAAACCTCAAGAGGAGGACCAACCCCTAATCTATGTGATGACAGTCACTGCCAAAACC	930
Qy	885	AGCGAGGAGATGTACAAGACCATCCAGTCCCTGGCCCAAGGCCCTCAGGAACGCATACCTGC	944
Db	931	AGCGAGGAGATGTACAAGACCATCCAGTCCCTGGCCCAAGGCCCTCAGGAAGGCATACCTGC	990
Qy	945	TTGTGCGCTATGAGAGACCTGGCTCGAGCCCTGTGGCCCAAGCTTCCGAAATGTATGAAT	1004
Db	991	TTGTGCGCTATGAGAGACCTGGCTCGAGCCCTGTGGCCCAAGCTTCCGAAATGTATGAAT	1050
Qy	1005	TCGTGGGATTTGGAAATTTCTGGCCCATCTTCAGACTGGGTGCATTAATCATCCGAGCA	1064
Db	1051	TCGTGGGATTTGGAAATTTCTGGCCCATCTTCAGACTGGGTGCATTAATCATCCGAGCA	1110
Qy	1065	AGGCGATGGGTGACACAGCTTTCCACACAATATGCAGAGGATGCCCTTAATGTGCCAGG	1124
Db	1111	AGGCGATGGGTGACACAGCTTTCCACACAATATGCAGAGGATGCCCTTAATGTGCCAGG	1170
Qy	1125	CTTGGCGCTGTCTTTGGCCCTATGAAAAAGTTTCTCGACTTCAGAAAAGCCTGTGGCGATG	1184
Db	1171	CTTGGCGCTGTCTTTGGCCCTATGAAAAAGTTTCTCGACTTCAGAAAAGCCTGTGGCGATG	1230
Qy	1185	CCATGAATTTCTGTGGGCTACCGGCACAGTCATGTGAACAGAAACAGAAACCTGTGTGC	1244
Db	1231	CCATGAATTTCTGTGGGCTACCGGCACAGTCATGTGAACAGAAACAGAAACCTGTGTGC	1290
Qy	1245	TGATATCTTCTGTCTACCTGAGACTGCTCCCTGAGCAATCCACTAAGAGGGTGTGAAGAGCT	1304
Db	1291	TGATATCTTCTGTCTACCTGAGACTGCTCCCTGAGCAATCCACTAAGAGGGTGTGAAGAGCT	1350
Qy	1305	TTTGTGTCCACCTGGTGTACAGCTCTAAGTACATTTCTCTGAATATGCTTGTGAGCTTGCTTAC	1364
Db	1351	TTTGTGTCCACCTGGTGTACAGCTCTAAGTACATTTCTCTGAATATGCTTGTGAGCTTGCTTAC	1410
Qy	1365	ATCTCTAGAGCTTTAACTACATGTCTGGGGATACACTGAGTGTGATGTTGTGCCACAC	1424
Db	1411	ATCTCTAGAGCTTTAACTACATGTCTGGGGATACACTGAGTGTGATGTTGTGCCACAC	1470
Qy	1425	GTGTCTCAAGCAGAAGACTTTTGTGTCCATGCTTGTCTAGAAAACAGACTGGGGAAAC	1484
Db	1471	GTGTCTCAAGCAGAAGACTTTTGTGTCCATGCTTGTCTAGAAAACAGACTGGGGAAAC	1530
Qy	1485	TTTATGTAGACAGCATATCCACACAGTGAACAGGGTATTTGCTCTTCTTTCTTTCTATC	1544
Db	1531	TTTATGTAGACAGCATATCCACACAGTGAACAGGGTATTTGCTCTTCTTTCTTTCTATC	1590

OY 1545 TTCCGTCTGGGAGACTTTCAGAGACTTGTGGCTGGAGGCTATTAAAGCAGCAGAC 1604
 |||||||
 DB 1591 TTCCGTCTGGGAGACTTTCAGAGACTTGTGGCTGGAGGCTATTAAAGCAGCAGAC 1650
 |||||||
 OY 1605 TATCAGTGGAAATGATTCATTAACCTCCCTGTCACATCTTGGCCAAATGGGAAATGATC 1664
 |||||||
 DB 1651 TATCAGTGGAAATGATTCATTAACCTCCCTGTCACATCTTGGCCAAATGGGAAATGATC 1710
 |||||||
 OY 1665 TTTCACCAAGAGCTCAGCAGCATTTTTCACAGAGATGC 1703
 |||||||
 DB 1711 TTTCACCAAGAGCTCAGCAGCATTTTTCACAGAGATGC 1749
 |||||||

RESULT 13

LOCUS BD124832 877 bp DNA linear PAT 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD124832
 VERSION BD124832.1 GI:23219777
 KEYWORDS JP 2002017375-A/263
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
 Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
 Koga,H.
 Primer for synthesizing full-length cDNA and use thereof
 Patent: JP 2002017375-A 263 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/263

COMMENT
 TITLE
 JOURNAL
 OS Homo sapiens (human)
 PN JP 2002017375-A/263
 PD 22-JAN-2002 JP 2000253172
 PE 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TETSUJI OTSUKI,HISASHI KOGA
 PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
 10,
 PC C12P21/02,C12O1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key

FT source 1..877 /organism='Homo sapiens (human)'.
 Location/Qualifiers
 1..877
 /organism='Homo sapiens'
 /mol_type='genomic DNA'
 /db_xref='taxon:9606'
 FT source 1..877 /organism='Homo sapiens (human)'.
 Location/Qualifiers
 1..877
 /organism='Homo sapiens'
 /mol_type='genomic DNA'
 /db_xref='taxon:9606'

BASE COUNT 183 a 266 c 221 g 203 t 4 others
 ORIGIN

Query Match 34.0%; Score 695; DB 6; Length 877;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 106 CAAAGTCTTCACCTTCAGACAAATGCTACTGCTTAAATAAATGAAGCTCCTGCTTCT 165
 |||||||
 DB 47 CAAAGTCTTCACCTTCAGACAAATGCTACTGCTTAAATAAATGAAGCTCCTGCTTCT 106
 |||||||
 OY 166 GGTTCACAGATGGCCATCTTGGCTATTCTTCCACATGTACAGCCACAATCAGCTC 225
 |||||||
 DB 107 GGTTCACAGATGGCCATCTTGGCTATTCTTCCACATGTACAGCCACAATCAGCTC 166
 |||||||
 OY 226 CCGTCTATGAAGGACAGCCCGAGCGCATGACAGTGTCTGCTCTTCTTCTGCGCTC 285
 |||||||
 DB 167 CCGTCTATGAAGGACAGCCCGAGCGCATGACAGTGTCTGCTCTTCTTCTGCGCTC 226
 |||||||

FEATURES
 source 1..877 /organism='Homo sapiens (human)'.
 Location/Qualifiers
 1..877
 /organism='Homo sapiens (human)'.
 Location/Qualifiers

OY 286 TGGCTCTTCTTTTGGGGAGACTTTTGGGAGACACCAGATGTTTCTACCTGATGA 345
 |||||||
 DB 227 TGGCTCTTCTTTTGGGGAGACTTTTGGGAGACACCAGATGTTTCTACCTGATGA 286
 |||||||
 OY 346 GCCCGCCGAGCAGTGTGATGATACCTTTCACAGAGACCCGCTGATGTCATGATGC 405
 |||||||
 DB 287 GCCCGCCGAGCAGTGTGATGATACCTTTCACAGAGACCCGCTGATGTCATGATGC 346
 |||||||
 OY 406 TGTGGGATCTGATACGGCCGCTTCTTGTGACATGAGCGTCTTATGCTTACAT 465
 |||||||
 DB 347 TGTGGGATCTGATACGGCCGCTTCTTGTGACATGAGCGTCTTATGCTTACAT 406
 |||||||
 OY 466 GGAACCTGCTCCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAACACCGGGCCCTGTG 525
 |||||||
 DB 407 GGAACCTGCTCCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAACACCGGGCCCTGTG 466
 |||||||
 OY 526 TTTCGACCTCCTGCTGACATCATCCCAAGATGAATATCCCGCGGCTCATGTCAG 585
 |||||||
 DB 467 TTTCGACCTCCTGCTGACATCATCCCAAGATGAATATCCCGCGGCTCATGTCAG 526
 |||||||
 OY 586 GCTCCTGTGACGTACAACAGCCCTTGTGAGGTGTGAGAGAGCCCTGCTCTACAGCA 645
 |||||||
 DB 527 GCTCCTGTGACGTACAACAGCCCTTGTGAGGTGTGAGAGAGCCCTGCTCTACAGCA 586
 |||||||
 OY 646 CGTGAGTCAAGAGAGTGGCGCTTCTTCAACCTGACAGTCCCTTACCGGCTGTAAGA 705
 |||||||
 DB 587 CGTGAGTCAAGAGAGTGGCGCTTCTTCAACCTGACAGTCCCTTACCGGCTGTAAGA 646
 |||||||
 OY 706 CCCCTCCCTCAACCTGATATGTCACCTGCTCGGAGACCCCGGCGGTTCGCTTC 765
 |||||||
 DB 647 CCCCTCCCTCAACCTGATATGTCACCTGCTCGGAGACCCCGGCGGTTCGCTTC 706
 |||||||
 OY 766 CCGAGAGCAGCAAGAGAGATCTCATATTGACA 800
 |||||||
 DB 707 CCGAGAGCAGCAAGAGAGATCTCATATTGACA 741
 |||||||

RESULT 14

LOCUS BD126950 877 bp DNA linear PAT 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD126950
 VERSION BD126950.1 GI:23221895
 KEYWORDS JP 2002017375-A/2381.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
 Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
 Koga,H.
 Primer for synthesizing full-length cDNA and use thereof
 Patent: JP 2002017375-A 2381 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/2381

COMMENT
 TITLE
 JOURNAL
 OS Homo sapiens (human)
 PN JP 2002017375-A/2381
 PD 22-JAN-2002 JP 2000253172
 PE 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TETSUJI OTSUKI,HISASHI KOGA
 PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
 10,
 PC C12P21/02,C12O1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key

FT source 1..877 /organism='Homo sapiens (human)'.
 Location/Qualifiers
 1..877
 /organism='Homo sapiens (human)'.
 Location/Qualifiers

source 1.877
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 183 a 266 c 221 g 203 t 4 others
 ORIGIN

Query Match 34.0%; Score 695; DB 6; Length 877;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 CAGGCTTCACATTCAGCAATGCTACTGCTAAATAAATCAACCTCTGCTGTTCT 165
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 47 CAGGCTTCACATTCAGCAATGCTACTGCTAAATAAATCAACCTCTGCTGTTCT 106
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 166 GGTTCACAGATGGCCATCTGGCTATCTTCCACATGTACAGCACAATCAGCTC 225
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 107 GGTTCACAGATGGCCATCTGGCTATCTTCCACATGTACAGCACAATCAGCTC 166
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 226 CCTGTATGAAAGGACAGCCGAGCGCATGACGTGCTGTCTCTCTGCGGCTC 285
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 167 CCTGTATGAAAGGACAGCCGAGCGCATGACGTGCTGTCTCTCTGCGGCTC 226
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 286 TGCGCTCTCTTGTGGGCGAGCTTTTGGGCGACACCGAGATGTTTCTACTGATGA 345
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 227 TGCGCTCTCTTGTGGGCGAGCTTTTGGGCGACACCGAGATGTTTCTACTGATGA 286
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 346 GCCCGCTGCGACAGTGTGATGACCTTCAAGCAGACACCGCTGATGCTGACATGGC 405
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 287 GCCCGCTGCGACAGTGTGATGACCTTCAAGCAGACACCGCTGATGCTGACATGGC 346
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 406 TGTGCGGATCTGATACGGGCGCTCTTGTGCGACATGAGCGCTTTGATGCTACAT 465
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 347 TGTGCGGATCTGATACGGGCGCTCTTGTGCGACATGAGCGCTTTGATGCTACAT 406
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 466 GGAACCTGATCCCGGAGACAGTCCAGCTCTTCAAGTGGGAGAAAGCGGCGCTGTG 525
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 407 GGAACCTGATCCCGGAGAGACAGCTCTTCAAGTGGGAGAAAGCGGCGCTGTG 466
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 526 TTCTGACCTGCTGTGACATCATCCACAGATGAATATCCCGGCGCTGACTGACG 585
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 467 TTCTGACCTGCTGTGACATCATCCACAGATGAATATCCCGGCGCTGACTGACG 526
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 586 GCTCCTGTCAGTCAACACACCCCTTTAGAGTGTGGAGAAAGCGCTGCTCTACAGCA 645
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 527 GCTCCTGTCAGTCAACACACCCCTTTAGAGTGTGGAGAAAGCGCTGCTCTACAGCA 586
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 646 CGTGTGTCAGAGAGAGTGGCTTCTTCAACCTGACAGTCCCTACCGCTGTGAAGA 705
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 587 CGTGTGTCAGAGAGAGTGGCTTCTTCAACCTGACAGTCCCTACCGCTGTGAAGA 646
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 706 CCCCTCCCTCAACCTGATATGCTGACCTGTGCGGACCCCGGCGCTGTTCGTTTC 765
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 647 CCCCTCCCTCAACCTGATATGCTGACCTGTGCGGACCCCGGCGCTGTTCGTTTC 706
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 766 CCGAGAACGACAAAGGAGATCTCATGATTGACA 800
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 707 CCGAGAACGACAAAGGAGATCTCATGATTGACA 741
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 15
 AX381256 517 bp DNA linear PAT 18-MAR-2002
 LOCUS AX381256
 DEFINITION Sequence 194 from Patent WO212280.
 ACCESSION AX381256
 VERSION AX381256.1 GI:19576075
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 1 Pyle, R.A., Xu, J. and Secrist, H.

TITLE Compositions and methods for the therapy and diagnosis of colon cancer
 JOURNAL Patent: WO 0212280-A 194 14-FEB-2002;
 CORIXA CORPORATION (US)
 FEATURES
 source 1.517
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 121 a 130 c 124 g 142 t
 ORIGIN

Query Match 19.7%; Score 403; DB 6; Length 517;
 Best Local Similarity 100.0%; Pred. No. 3.5e-228;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GGCTTGGCCACCGGAGTGTAGCCTCTGATCACTTCTCTAATGCTTGTAGCCTTGC 1360
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 13 GGCTTGGCCACCGGAGTGTAGCCTCTGATCACTTCTCTAATGCTTGTAGCCTTGC 72
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1361 CTACATCTGAGCCTTAATCAATGCTGTGGGATTCACACTGAGTGTGATGCTC 1420
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 73 CTACATCTGAGCCTTAATCAATGCTGTGGGATTCACACTGAGTGTGATGCTC 132
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1421 ACACGTGTCAGACAGAGACTTTTGTGTCCATGCTGTGTAGAAAAACAGACTGGG 1480
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 133 ACACGTGTCAGACAGAGACTTTTGTGTCCATGCTGTGTAGAAAAACAGACTGGG 192
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1481 AACCTTATGAGACACATCCACAGTGAAGAGGATGCTCTCTTCTTCTT 1540
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 193 AACCTTATGAGACACATCCACAGTGAAGAGGATGCTCTCTTCTTCTTCTT 252
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1541 GATCTTCTGCTGTGGGACAGTTCAGAGACTTTGTGGCTGAGGCGCTATTAGACAGAC 1600
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 253 GATCTTCTGCTGTGGGACAGTTCAGAGACTTTGTGGCTGAGGCGCTATTAGACAGAC 312
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1601 ACAGTATCAGTGAATGATTCATTAACCTCCCTGTCCACATCTTCCCATAGGGGATG 1660
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 313 ACAGTATCAGTGAATGATTCATTAACCTCCCTGTCCACATCTTCCCATAGGGGATG 372
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1661 GATCTTTCACAAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 373 GATCTTTCACAAAGAGCTCACAGCATTTTCCACAGAGATGC 415
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Search completed: August 15, 2003, 03:08:35
 Job time : 7285 secs

XX 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 XX
 PS Claim 8; SEQ ID NO 2816; 1380bp + sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5' - and 3' - ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a full length
 CC human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 XX
 SQ Sequence 1979 BP; 451 A; 542 C; 474 G; 512 T; 0 other;
 XX
 Query Match 75.7%; Score 1547; DB 22; Length 1979;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 106 CAGGCTCTTCAGCTCAGCAGATGCTACTGCTTAAATAATGAAAGCTGCTGTTCT 165
 DB 47 CAGGCTCTTCAGCTCAGCAGATGCTACTGCTTAAATAATGAAAGCTGCTGTTCT 106
 QY 166 GGTTCCTCCAGATGGGCGATCTGGCTATCTTCCAGATGATGAGCAGCAATCAGCTC 225
 DB 107 GGTTCCTCCAGATGGGCGATCTGGCTATCTTCCAGATGATGAGCAGCAATCAGCTC 166
 QY 226 CCGTCTATGAAGGACACACCCGAGGCGATGCACTGCTGTTCTTCCAGGCGCTC 285
 DB 167 CCGTCTATGAAGGACACACCCGAGGCGATGCACTGCTGTTCTTCCAGGCGCTC 226
 QY 286 TGGCTCTTCTTTTGGGGGAGCTTTTGGGAGACCCAGATGTTTCTTACCTGATGA 345
 DB 227 TGGCTCTTCTTTTGGGGGAGCTTTTGGGAGACCCAGATGTTTCTTACCTGATGA 286
 QY 346 GCCCGCTGCGACAGTGTGATGACCTTCAAGCAGAGACACCGCTGATCTGACATGCG 405
 DB 287 GCCCGCTGCGACAGTGTGATGACCTTCAAGCAGAGACACCGCTGATCTGACATGCG 346
 QY 406 TGTGCGGAGATGATACGGGCGCTCTTCTTGTGCGACATGAGCGTCTTGTATGCTTACAT 465
 DB 347 TGTGCGGAGATGATACGGGCGCTCTTCTTGTGCGACATGAGCGTCTTGTATGCTTACAT 406
 QY 466 GGAACCTGTCCTCCCGAGACATCCAGCCTCTTTCAGTGGGAGAACAGCCGGCCCTGTG 525
 DB 407 GGAACCTGTCCTCCCGAGACATCCAGCCTCTTTCAGTGGGAGAACAGCCGGCCCTGTG 466
 QY 526 TTCTGACACTGCTGTGACATCATCCACAAATGAAATCATCCCGGGGCTCAGTGCAG 585
 DB 467 TTCTGACACTGCTGTGACATCATCCACAAATGAAATCATCCCGGGGCTCAGTGCAG 526
 QY 586 GCTCTGTGACATCAACAGCCCTTGTGAGGTGTGAGAAAGGCTGCGCTCTTACAGCCA 645
 DB 527 GCTCTGTGACATCAACAGCCCTTGTGAGGTGTGAGAAAGGCTGCGCTCTTACAGCCA 586
 QY 646 GCTGTGCTCAAGAGGAGTGTGCTTCAACCTGCAAGTCCCTTACCCGCTGCTGAAGA 705
 DB 587 GCTGTGCTCAAGAGGAGTGTGCTTCAACCTGCAAGTCCCTTACCCGCTGCTGAAGA 646
 QY 706 CCCCTCCCTCAACCTGATATGTCGACCTGTGCGGAGCCCGGGGCGCTGTCCGCTTC 765
 DB 647 CCCCTCCCTCAACCTGATATGTCGACCTGTGCGGAGCCCGGGGCGCTGTCCGCTTC 706
 QY 766 CCGAAGACGACAAAGGAGATCTCATGATTGACAGTGCATGTCATGTCGAGGAGCATGA 825
 DB 707 CCGAAGACGACAAAGGAGATCTCATGATTGACAGTGCATGTCATGTCGAGGAGCATGA 766

QY 826 GCAAAAACCTCAGAGAGGAGACCACCTACTATGTGATGACAGTATCTGCCAAGCCA 885
 DB 767 GCAAAAACCTCAGAGAGGAGACCACCTACTATGTGATGACAGTATCTGCCAAGCCA 826
 QY 886 GCTGGAGATCTACAGAACCATTCAGTCTTGGCCAAAGGCCCTGACAGAACCTTCTGCT 945
 DB 827 GCTGGAGATCTACAGAACCATTCAGTCTTGGCCAAAGGCCCTGACAGAACCTTCTGCT 886
 QY 946 TGTGGCTATGAGAGGAGCTGGCTGAGCCCTGAGCCAGACTTCCGGAATGATGAAT 1005
 DB 887 TGTGGCTATGAGAGGAGCTGGCTGAGCCCTGAGCCAGACTTCCGGAATGATGAAT 946
 QY 1006 CGTGGATGGAATTTCTTCCCATCTTCAAGCTGCTGATGATATACATCACCAGGCAA 1065
 DB 947 CGTGGATGGAATTTCTTCCCATCTTCAAGCTGCTGATGATATACATCACCAGGCAA 1006
 QY 1066 GGGCATGGGTGACACAGCTTTTCCACAAATGCGAGGATGCCCTTAATGTCGCCAGGC 1125
 DB 1007 GGGCATGGGTGACACAGCTTTTCCACAAATGCGAGGATGCCCTTAATGTCGCCAGGC 1066
 QY 1126 TTGGCGCTGCTTGGCCCTATGAAAGGTTTCTGACTTCAGAAAGCCGTGGGAGTGC 1185
 DB 1067 TTGGCGCTGCTTGGCCCTATGAAAGGTTTCTGACTTCAGAAAGCCGTGGGAGTGC 1126
 QY 1186 CATGAATTTGCTGGGCTACCGCCAGCTCAGATCTGAAGAAGACAGAGAAACCTGTGCT 1245
 DB 1127 CATGAATTTGCTGGGCTACCGCCAGCTCAGATCTGAAGAAGACAGAGAAACCTGTGCT 1186
 QY 1246 GATCTTCTGTCTACTGAGACGTGCTCCCTAGAGAAATCCACTAAGAGGTTGAGAAAGCTT 1305
 DB 1187 GATCTTCTGTCTACTGAGACGTGCTCCCTAGAGAAATCCACTAAGAGGTTGAGAAAGCTT 1246
 QY 1306 TGTGCGACACCTGGTGTGACCTCAGCTCAGTCACTTCTGTAATGCTTGTGACCTTGTCTACA 1365
 DB 1247 TGTGCGACACCTGGTGTGACCTCAGCTCAGTCACTTCTGTAATGCTTGTGACCTTGTCTACA 1306
 QY 1366 TCTCTGAGCTTAACTAATCTGTCTGGGTATCACAGTATGATGATGTTGTCCACAGC 1425
 DB 1307 TCTCTGAGCTTAACTAATCTGTCTGGGTATCACAGTATGATGATGTTGTCCACAGC 1366
 QY 1426 TGTCTAAGCAGAGAGCTTGTGTGTCATGCTGTGTGTAAGAAACAGACTGGGAAAGCT 1485
 DB 1367 TGTCTAAGCAGAGAGCTTGTGTGTCATGCTGTGTGTAAGAAACAGACTGGGAAAGCT 1426
 QY 1486 TATGTGAGCAGACATCCACAGTGAACAGAGGATATGCTCTTCTTCTTCTGATCT 1545
 DB 1427 TATGTGAGCAGACATCCACAGTGAACAGAGGATATGCTCTTCTTCTTCTGATCT 1486
 QY 1546 TCTCTGTCTGGGAGACTTCAAGAGACTTGTGGCTGGAGAGCTTATTAACAGACACAT 1605
 DB 1487 TCTCTGTCTGGGAGACTTCAAGAGACTTGTGGCTGGAGAGCTTATTAACAGACACAT 1546
 QY 1606 ATCAGTGAATTTGATCATTAACCTCCCTGTCACATCTTGGCCAAATGGGAGATGATCT 1665
 DB 1547 ATCAGTGAATTTGATCATTAACCTCCCTGTCACATCTTGGCCAAATGGGAGATGATCT 1606
 QY 1666 TTACCAAAAGAGCTCACAGCAATTTTCCACAGAGATGC 1703
 DB 1607 TTACCAAAAGAGCTCACAGCAATTTTCCACAGAGATGC 1644

RESULT 2
 AA220792
 ID AA220792 standard; DNA; 2032 BP.
 XX
 XX AA220792;
 AC
 XX
 DT 08-DEC-1999 (first entry)
 XX
 XX Human glycosyl sulfotransferase-3 coding sequence.
 DE Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
 KW selectin binding interaction; inflammation; lymphocyte homing; human;

KW secondary lymph organ; ss.
XX Homo sapiens.
XX W09949018-A1.
XX 30-SEP-1999.
XX 26-FEB-1999; 99WO-US04316.
XX 20-MAR-1998; 98US-0045284.
XX 12-NOV-1998; 98US-0190911.
XX (REGC) UNIV CALIFORNIA.
XX (SYNT) SYNTAX USA INC.
XX Blstrup A, Rosen SD, Tangemann K, Hemmerlich S;
XX WPI; 1999-580442/49.
XX P-PSDB; AAY39918.
XX Human and murine glycosyl sulfotransferase 3 and related
XX polynucleotides
XX
XX Claim 4; Fig 1; 59pp; English.
XX
XX This sequence encodes the human glycosyl sulfotransferase-3 (GST-3) of
XX the invention. The nucleic acid sequences, probes and primers derived
XX from these, proteins and antibodies are useful in detecting homologues.
XX The sequences, antibodies and methods are useful in the diagnosis and
XX treatment of diseases associated with selectin binding interactions,
XX including conditions associated with or resulting from the homing of
XX leukocytes to sites of inflammation and the normal homing of lymphocytes
XX to secondary lymph organs.
XX
XX Sequence 2032 BP; 468 A; 569 C; 490 G; 505 T; 0 other;
SQ
Query Match 75.7%; Score 1547; DB 20; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 106 CAAAGCTCTCCATTCAGACACATGCTACTGCTTAAATAAAGAACTGCTGCTTTCT 165
DB 160 CAAAGCTCTCCATTCAGACACATGCTACTGCTTAAATAAAGAACTGCTGCTTTCT 219
QY 166 GGTTCACAGATGGCCATCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 225
DB 220 GGTTCACAGATGGCCATCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 279
QY 226 CCTGCTATGAGGACACAGCCGAGCGATGACGTGCTGCTTCTTCTTCTTCTTCTTCTTCT 285
DB 280 CCTGCTATGAGGACACAGCCGAGCGATGACGTGCTGCTTCTTCTTCTTCTTCTTCTTCT 339
QY 286 TGCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 345
DB 340 TGCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 399
QY 346 GCCCGCTGACGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 405
DB 400 GCCCGCTGACGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 459
QY 406 TGTGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 465
DB 460 TGTGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 519
QY 466 GGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 525
DB 520 GGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 579
QY 526 TTCGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 585
DB 580 TTCGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 639

QY 586 GCTCCTGTGACATCAACACCCCTTTGAGGTGTGAGAGAGGCTGCGCTCTCTACAGCA 645
DB 640 GCTCCTGTGACATCAACACCCCTTTGAGGTGTGAGAGAGGCTGCGCTCTCTACAGCA 699
QY 646 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGCAAGTCCCTTACCCGCTGTAAGA 705
DB 700 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGCAAGTCCCTTACCCGCTGTAAGA 759
QY 706 CCCCTCCCTCAACCTGCAATGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 765
DB 760 CCCCTCCCTCAACCTGCAATGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 819
QY 766 CCGAGAACGCAAAAGGAGATCTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 825
DB 820 CCGAGAACGCAAAAGGAGATCTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 879
QY 826 GCAAAACTCAAGAGAGAGACCAACCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 885
DB 880 GCAAAACTCAAGAGAGAGACCAACCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 939
QY 886 GCTGAGATCTCAAGAGACATCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 945
DB 940 GCTGAGATCTCAAGAGACATCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 999
QY 946 TGTGCGCTATGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1005
DB 1000 TGTGCGCTATGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1059
QY 1006 CGTGGATTTGAAATTTCTTGGCCCATCTTCAAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1065
DB 1060 CGTGGATTTGAAATTTCTTGGCCCATCTTCAAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1119
QY 1066 GGGCATGGGTGACCAAGCTTTTCCACAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1125
DB 1120 GGGCATGGGTGACCAAGCTTTTCCACAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1179
QY 1126 TTGGCGCTGCTTCTTGGCCCTATGAAAGGTTTCTGATGATGAAAGGCTGCTGCTGCTGCTGCT 1185
DB 1180 TTGGCGCTGCTTCTTGGCCCTATGAAAGGTTTCTGATGATGAAAGGCTGCTGCTGCTGCTGCT 1239
QY 1186 CATGATTTGCTGGGCTACCGCCACGTGATCTGACACAGACAGACAGACAGACAGACAGACAGAC 1245
DB 1240 CATGATTTGCTGGGCTACCGCCACGTGATCTGACACAGACAGACAGACAGACAGACAGACAGAC 1299
QY 1246 GGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1305
DB 1300 GGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1359
QY 1306 TGCTGCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1365
DB 1360 TGCTGCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1419
QY 1366 TCTCTGAGCCTTAATACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1425
DB 1420 TCTCTGAGCCTTAATACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1479
QY 1426 TGCTCAACGAGAGACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1485
DB 1480 TGCTCAACGAGAGACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1539
QY 1486 TATGTGAGCAGACATCCACAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1545
DB 1540 TATGTGAGCAGACATCCACAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1599
QY 1546 TCTGTGCTGGGACGCTTCAAGACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1605
DB 1600 TCTGTGCTGGGACGCTTCAAGACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1659
QY 1606 ATCAGTGAATTTGATTAACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1665
DB 1660 ATCAGTGAATTTGATTAACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1719
QY 1666 TTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 1703

Db 1720 TTCACCAAGAGCTCACAGCATTTTCACAGAGATGC 1757
 |||
 RESULT 3
 ID AAS16947 standard; cDNA; 1333 BP.
 AAS16947;
 12-MAR-2002 (first entry)
 Human L-selectin sulfotransferase-2 (LST-2) cDNA.
 Human: beta1,3GNT; beta1,3-N-acetylglucosaminyltransferase; MECA-79; ss;
 L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
 ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;
 allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;
 delayed-type hypersensitivity reaction; hyperplastic thymus; antileuker;
 antiinflammatory; antipsoriatic; antidiabetic; dermatological;
 antiallergic.
 Homo sapiens.
 Key Location/Qualifiers
 CDS 111..1253
 /tag= a
 /product= "Human LST-2"
 WO200185177-A1.
 15-NOV-2001.
 10-MAY-2001; 2001WO-US15452.
 11-MAY-2000; 2000US-0569320.
 (BURN-) BURHAM INST.
 Fukuda M, Yeh J, Hiraoaka N;
 MPI; 2002-075226/10.
 P-PSDB; AAU11274.
 New enzyme, useful for modifying acceptor molecule, comprises an
 isolated L-selectin sulfotransferase-2 that directs expression of
 L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
 intestinal GlcNAc 6-sulfotransferase
 Claim 19; Fig 4; 98pp; English.
 The present invention provides a method of modifying an acceptor molecule
 by contacting the acceptor with an isolated
 beta1,3-N-acetylglucosaminyltransferase (beta1,3GNT) or an active
 fragment, where beta1,3GNT directs expression of a MECA-79 antigen. The
 invention also provides a method of treating or preventing an
 L-selectin-mediated condition by reducing the expression or activity of a
 beta1,3GNT that directs expression of a MECA-79 antigen. This can be done
 by administering to the subject an oligosaccharide L-selectin antagonist
 that inhibits binding of L-selectin to a MECA-79 antigen, for example by
 administering antibody material that specifically binds beta1,3GNT,
 and/or a beta1,3GNT antisense nucleic acid molecule. L-selectin
 sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.
 Alternatively, the expression or activity of LST-2 or its active
 fragment can be reduced in combination with reducing the expression or
 activity of beta1,3GNT. The method is useful for treating L-selectin
 mediated conditions such as Crohn's disease and ulcerative colitis,
 inflammatory disorders of the skin such as allergic contact dermatitis,
 psoriasis and Lichen planus, lymphomas, chronic pneumonia, delayed-type
 hypersensitivity reactions, diabetes and hyperplastic thymus. This
 sequence represents cDNA encoding human LST-2.
 Sequence 1333 BP; 286 A; 393 C; 336 G; 318 T; 0 other;

Query Match 61.8%; Score 1262; DB 24; Length 1333;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 107 AAGCTTCCACCTTCACAGCAGATGCTACTGCTTAAATAAATGAAGCTCTGCTGTTCTG 166
 72 AAGGCTCTCCACTTCACAGCAGATGCTACTGCTTAAATAAATGAAGCTCTGCTGTTCTG 131
 167 GTTCCAGATGAGGCTCTGCTTAAATGCTTCAATGATGACAGCAGCAATGATGCTCTC 226
 132 GTTCCAGATGAGGCTCTGCTTAAATGCTTCAATGATGACAGCAGCAATGATGCTCTC 191
 227 CTGCTATGAGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 286
 192 CTGCTATGAGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 251
 287 GGTCTCTTCTTTTGTGGGACGCTTTTGGGACAGCAGCAGCAGCAGCAGCAGCAGCAG 346
 252 GGTCTCTTCTTTTGTGGGACGCTTTTGGGACAGCAGCAGCAGCAGCAGCAGCAGCAG 311
 347 CCCGCTGACAGCTGTGATGACCTTCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 406
 312 CCCGCTGACAGCTGTGATGACCTTCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 371
 407 GTGCGGAGTGTGATGACGCGCTCTTCTGTGCGACATGACGCTTTGATGCTTACATG 466
 372 GTGCGGAGTGTGATGACGCGCTCTTCTGTGCGACATGACGCTTTGATGCTTACATG 431
 467 GAACCTGCTCCCGGAGAGAGCTGACGCTCTTCAAGGAGAGAGAGAGAGAGAGAGAGAG 526
 432 GAACCTGCTCCCGGAGAGAGCTGACGCTCTTCAAGGAGAGAGAGAGAGAGAGAGAGAG 491
 527 TCTGACAGCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 586
 492 TCTGACAGCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 551
 587 CTCTGCTGAGTGAACAGCCCTTTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 646
 552 CTCTGCTGAGTGAACAGCCCTTTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 611
 647 GTGCTGCTCAAG 706
 612 GTGCTGCTCAAG 671
 707 CCTCTCCCTCAACCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 766
 672 CCTCTCCCTCAACCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 731
 767 CGAGAACGCAAAAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 826
 732 CGAGAACGCAAAAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 791
 827 CAATAAATCAAG 886
 792 CAATAAATCAAG 851
 887 CTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 946
 852 CTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 911
 947 GTGCGTATGAG 1006
 912 GTGCGTATGAG 971
 1007 GTGCGATGAG 1066
 972 GTGCGATGAG 1031
 1067 GGCATGGGTGACACAGCTTTCCACAAATGCGAGGAGATGCTTATGCTTCCAGGCT 1126
 1032 GGCATGGGTGACACAGCTTTCCACAAATGCGAGGAGATGCTTATGCTTCCAGGCT 1091

QY 1127 TGCGGCTGCTCTTGGCCCTATGAAAGTTTCTGACTTCGAAAGCCCTGTGGGATGCC 1186
 DB 1092 TGGCGCTGCTCTTGGCCCTATGAAAGTTTCTGACTTCGAAAGCCCTGTGGGATGCC 1151
 QY 1187 ATGAATTTGCTGGGCTACCCGACGATCTGAACAGACAGAAACCTTGTCTG 1246
 DB 1152 ATGAATTTGCTGGGCTACCCGACGATCTGAACAGACAGAAACCTTGTCTG 1211
 QY 1247 GATCTCTGCTACCTGAGCTGCTGAGCAATTCACCTAGAGGGGTGAGAGGCTTT 1306
 DB 1212 GATCTCTGCTACCTGAGCTGCTGAGCAATTCACCTAGAGGGGTGAGAGGCTTT 1271
 QY 1307 GCTGCCACCTGCTGCTGACCTCAGTCACTTCTCTGATGCTTGTGAGCCCTGCTACAT 1366
 DB 1272 GCTGCCACCTGCTGCTGACCTCAGTCACTTCTCTGATGCTTGTGAGCCCTGCTACAT 1331
 QY 1367 CT 1368
 DB 1332 CT 1333

RESULT 4
 AA294211
 ID AA294211 standard; cDNA; 2065 BP.
 AC AA294211;
 XX 19-JUN-2000 (first entry)
 DE Human transferase TRNSFS-11 cDNA clone 2617407CBI.
 XX
 XX Transferase; TRNSFS-11; human; antitumor; cell proliferation;
 KM inflammation; gastrointestinal disorder; developmental disorder;
 KM genetic disorder; neurological disorder; reproductive disorder;
 KM smooth muscle disorder; immunological disorder; gene therapy;
 KM diagnosis; N-acetylglucosamine 6-O-sulfotransferase; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 174..1334
 FT /*tag= a
 XX
 PN MO200014251-A2.
 XX
 PD 16-MAR-2000.
 XX
 PF 09-SEP-1999; 99WO-US20989.
 XX
 PR 10-SEP-1998; 98US-0150657.
 PR 04-NOV-1998; 98US-0186779.
 PR 11-MAY-1999; 99US-0133642.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Tang YT, Corley NC, Guegler KJ, Baughn MR, Lai P, Yue H;
 PI Hillman JL, Azimzai Y;
 XX
 DR WPI: 2000-256996/22.
 DR P-PSDB: AAY79219.
 XX
 XX Human transferase proteins useful for preventing, diagnosing and
 PT treating cancers and developmental, gastrointestinal, genetic,
 PT immunological, neurological, reproductive and smooth muscle disorders -
 XX
 PS Claim 9; Page 104-105; 113pp; English.
 CC
 CC The present sequence is that of cDNA clone 2617407CBI encoding
 CC human transferase TRNSFS-11 (see AAY79219), 1 of 15 claimed human
 CC transferase proteins of the invention (see AAY79209-23). The clone
 CC was isolated from gall bladder cDNA library GBIANT01. TRNSFS-11
 CC is expressed in dermatologic and gastrointestinal tissues,
 CC especially those associated with inflammation and cell

CC proliferation. It shows homology to mouse N-acetylglucosamine
 CC 6-O-sulfotransferase. A polynucleotide comprising nucleotides
 CC 264-333 or 1272-1331 of the present sequence can be used as a DNA
 CC probe. The new human transferases and polynucleotides can be used
 CC in the diagnosis, prevention and treatment (including gene therapy
 CC and antisense therapy) of cancer, developmental disorders,
 CC gastrointestinal disorders, genetic disorders, immunological
 CC disorders, neurological disorders, reproductive disorders, and
 CC smooth muscle disorders.
 XX
 SO Sequence 2065 BP; 466 A; 575 C; 491 G; 533 T; 0 other.

Query Match 56.4%; Score 1153; DB 21; Length 2065;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 1593; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 106 CAAAGCTTCCACTTACAGCAGCAATGCTACGCTTAAAAAATGAAAGCTCTGCTGTTCT 165
 DB 152 CAAAGCTTCCACTTACAGCAGCAATGCTACGCTTAAAAAATGAAAGCTCTGCTGTTCT 211
 QY 166 GGTTCACAGATGGCCATCTTGGCTATTTCTTCACATGTACAGCCACAATCAGCTC 225
 DB 212 GGTTCACAGATGGCCATCTTGGCTATTTCTTCACATGTACAGCCACAATCAGCTC 271
 QY 226 CCTGCTATGAAAGCAGACGCCGAGCGATGACGTGCTGCTGCTTCTGCTGCGCTC 285
 DB 272 CCTGCTATGAAAGCAGACGCCGAGCGATGACGTGCTGCTGCTTCTGCTGCGCTC 331
 QY 286 TGCGCTCTCTTTTGTGGGGAGCTTTTGGGACACCCAGCATTTTCTACCTGATGA 345
 DB 332 TGCGCTCTCTTTTGTGGGGAGCTTTTGGGACACCCAGCATTTTCTACCTGATGA 391
 QY 346 GCCCGCTGACCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 405
 DB 392 GCCCGCTGACCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 451
 QY 406 TGTCGGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 465
 DB 452 TGTCGGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 511
 QY 466 GGAACCTGTGCTCCCGGAGACAGTCCAGCTCTTCACTGAGGAGAAACAGCGGCGCTGTG 525
 DB 512 GGAACCTGTGCTCCCGGAGACAGTCCAGCTCTTCACTGAGGAGAAACAGCGGCGCTGTG 571
 QY 526 TTCTGACACCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 585
 DB 572 TTCTGACACCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 630
 QY 586 GCTCCTGTGACGTCAGACAGCCCTTTG-AGGTGTGTGAGAAAGGCTGCGGCTCTACAGCC 644
 DB 631 GCTCCTGTGACGTCAGACAGCCCTTTGAAAGTTGTTGAGAAAGCTGCGGCTCTACAGCC 690
 QY 645 ACGTGTGCTCAGAGAGTGGCTTTTCAACTGACAGTCCCTTACCCGCTGTGAAG 704
 DB 691 ACGTGTGCTCAGAGAGTGGCTTTTCAACTGACAGTCCCTTACCCGCTGTGAAG 750
 QY 705 ACCCTCCCTCAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 764
 DB 751 ACCCTCCCTCAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 810
 QY 765 CCCGAGAACGCAAAAGGAGATCTCATGATGATGATGATGATGATGATGATGATGATGATG 824
 DB 811 CCCGAGAACGCAAAAGGAGATCTCATGATGATGATGATGATGATGATGATGATGATGATGATG 870
 QY 825 AGCAAAATCTCAAGAGAGAGACCAACCTCTACTGTGTGATGATGATGATGATGATGATG 884
 DB 871 AGCAAAATCTCAAGAGAGAGACCAACCTCTACTGTGTGATGATGATGATGATGATGATG 930
 QY 885 AGCTGAGATCTACAGAGCATCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 944
 DB 931 AGCTGAGATCTACAGAGCATCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 990
 QY 945 TTGTGCGCTATGAGACCTGCTGAGACCCCTGTGTGCGCAAGACTTCCCGAATGTATGAAT 1004

```

Db      991 TTTGTCGCTATGAGGAGCTGGCTCGAGCCCTGTGGCCAGACTTCCGAGATGTATGAAAT 1050
QY      1005 TCGTGGGATTTGAAATCTTGTCCCATTTCCAGACTGGGTGATTAATACATCACCAGGCA 1064
Db      1051 TCGTGGGATTTGAAATCTTGTCCCATTTCCAGACTGGGTGATTAATACATCACCAGGCA 1110
QY      1065 AGGGCATGGGTGACACCGCTTTCCACACAATGCCAGGATGCCCTTAATGTCTCCAGG 1124
Db      1111 AGGGCATGGGTGACACCGCTTTCCACACAATGCCAGGATGCCCTTAATGTCTCCAGG 1170
QY      1125 CTGGGCTGTGTCTTGGCCCTATGAAAGGTTCTGAGCTTGAAAGGCTGTGGGATG 1184
Db      1171 CTGGGCTGTGTCTTGGCCCTATGAAAGGTTCTGAGCTTGAAAGGCTGTGGGATG 1230
QY      1185 CCATGAAATTTGCTGGGCTACCGCCAGCTGAGATCTGACAGAAAGAGAAACCTGTTCG 1244
Db      1231 CCATGAAATTTGCTGGGCTACCGCCAGCTGAGATCTGACAGAAAGAGAAACCTGTTCG 1290
QY      1245 TGGATCTTCTGCTTACTGAGACTGTGCTCCAGCAATCCACTAAGAGGTTGAGAGGCT 1304
Db      1291 TGGATCTTCTGCTTACTGAGACTGTGCTCCAGCAATCCACTAAGAGGTTGAGAGGCT 1350
QY      1305 TTGCTGCACTGTGTGCTGAGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1364
Db      1351 TTGCTGCACTGTGTGCTGAGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1410
QY      1365 ATCTGAGGCTTAACTAATGCTGTGGGTATCAGACAGTGTGATGTTGTGCTCACAC 1424
Db      1411 ATCTGAGGCTTAACTAATGCTGTGGGTATCAGACAGTGTGATGTTGTGCTCACAC 1470
QY      1425 GTCCTGACAGAGAGAGCTTTGTGTCATGCTGTGTGTCAGAAACAGACGTGGGAGAC 1484
Db      1471 GTCCTGACAGAGAGAGCTTTGTGTCATGCTGTGTGTCAGAAACAGACGTGGGAGAC 1530
QY      1485 TATGTGAGCAGACATCCACAGTGAAGAAAGGATTTGCTCTTCTTCTTCTGATC 1544
Db      1531 TATGTGAGCAGACATCCACAGTGAAGAAAGGATTTGCTCTTCTTCTTCTGATC 1590
QY      1545 TTTCTGTCTGGGAGACTTCCAGAGACTTTGTGGCTGTGAGGCTTATTAAGCAGACAG 1604
Db      1591 TTTCTGTCTGGGAGACTTCCAGAGACTTTGTGGCTGTGAGGCTTATTAAGCAGACAG 1650
QY      1605 TATCAGTGAATTTGATCCATAAACCCTGTCATCATCTGCCCCAATGGGAGATGGATC 1664
Db      1651 TATCAGTGAATTTGATCCATAAACCCTGTCATCATCTGCCCCAATGGGAGATGGATC 1710
QY      1665 TTTACACAAAGAGCTCAGCAGATTTTCCACAGAGATGC 1703
Db      1711 TTTACACAAAGAGCTCAGCAGATTTTCCACAGAGATGC 1749

RESULT 5
AAK91803
ID      AAK91803 standard; cDNA; 877 BP.
XX
AC      AAK91803;
XX
DT      06-NOV-2001 (first entry)
XX
DE      Human cDNA 5'-end sequence, SEQ ID NO: 263.
XX
KW      Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS      Homo sapiens.
XX
PN      EP130094-A2.
XX
PD      05-SEP-2001.
XX
PF      07-JUL-2000; 2000BP-0114089.
XX
PR      08-JUL-1999; 99JP-0194486.

```

```

PR      11-JAN-2000; 2000BP-0118774.
PR      02-MAY-2000; 2000BP-0183765.
XX
XX      (HELI-) HELIX RES INST.
PI      Oka T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI      Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX      WPI; 2001-524255/58.
XX
PT      830 Primers useful for synthesizing full length cDNA clones and their
PT      use in genetic manipulation -
PS      Claim 2; SEQ ID NO 263; 1380bp + sequence listing; English.
XX
XX      The invention relates to primers for synthesizing full length cDNA
CC      clones. 830 cDNA molecules encoding a human protein have been
CC      isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC      molecules have been determined. Primers for synthesizing the full length
CC      cDNA are useful for clarifying the function of the protein encoded by
CC      the cDNA. The full length clones were obtained by construction of full
CC      length enriched cDNA libraries that were synthesised by the oligo-capping
CC      method. The primers enable the production of the full length cDNA easily
CC      without any special methods. The present sequence is the nucleotide
CC      sequence of the 5'-end of a cDNA provided in the invention.
CC      Note: The sequence data for this patent did not form part of the printed
XX      specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ      Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;
XX
Query Match      34.0%; Score 695; DB 22; Length 877;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY      106 CAAGTCTTCCACTTCCAGCACAATGCTACTGCTTAAATAAATGAACTCTGCTTTCT 165
Db      47 CAAGTCTTCCACTTCCAGCACAATGCTACTGCTTAAATAAATGAACTCTGCTTTCT 106
QY      166 GGTTCCTCCAGATGGCATCTTGCTATTTCTTCCACATGTACAGCCACATCAGCTC 225
Db      107 GGTTCCTCCAGATGGCATCTTGCTATTTCTTCCACATGTACAGCCACATCAGCTC 166
QY      226 CCGTCTATGAAGGACACCGGAGCGATGACAGCGTGGTCTTCTCTGGGGCTC 285
Db      167 CCGTCTATGAAGGACACCGGAGCGATGACAGCGTGGTCTTCTCTGGGGCTC 226
QY      286 TGGCTCTTCTTGTGTGGGAGAGCTTTTGGGAGCAGCCAGATGTTTCTTACTGATGA 345
Db      227 TGGCTCTTCTTGTGTGGGAGAGCTTTTGGGAGCAGCCAGATGTTTCTTACTGATGA 286
QY      346 GCCCGCTGGACAGTGTGATGACTTTCAGCAGACAGCAGCCGCTGGATCTGACATGCG 405
Db      287 GCCCGCTGGACAGTGTGATGACTTTCAGCAGACAGCAGCCGCTGGATCTGACATGCG 346
QY      406 TGTGGGGATGTGATACGGGGCGCTCTTCTTGTGCGCATGAGACGCTTTTGATGCCATAT 465
Db      347 TGTGGGGATGTGATACGGGGCGCTCTTCTTGTGCGCATGAGACGCTTTTGATGCCATAT 406
QY      466 GGAACCTGTGCTCCCGGAGAGAGTCCAGCTCTTTCAGTGGGAGAACAGCGGGGCTGTG 525
Db      407 GGAACCTGTGCTCCCGGAGAGAGTCCAGCTCTTTCAGTGGGAGAACAGCGGGGCTGTG 466
QY      526 TTTCTGACCTGCTGTGACATCATCCACAGATGAATCATCCCGGGGCTCACTGACG 585
Db      467 TTTCTGACCTGCTGTGACATCATCCACAGATGAATCATCCCGGGGCTCACTGACG 526
QY      586 GCTCTGTGCTCAACACAGCCCTTTGAGGTGTGAGAGAGGCTGCGCTCTTACAGCA 645
Db      527 GCTCTGTGCTCAACACAGCCCTTTGAGGTGTGAGAGAGGCTGCGCTCTTACAGCA 586
QY      646 CGTGGGTTCAGAGGAGGTGCGCTTCTCAACCTGACAGTCCCTTACCGGCTGTAAAGA 705
Db      587 CGTGGGTTCAGAGGAGGTGCGCTTCTCAACCTGACAGTCCCTTACCGGCTGTAAAGA 646

```


XX New isolated tumor colon polynucleotide and polypeptide, useful for the
 PT diagnosis, prevention and/or treatment of cancer, in particular colon
 PT cancer
 XX
 PS Claim 1: SEQ ID NO 2595; 266bp + Sequence Listing; English.
 CC
 CC The invention relates to a human colon tumour expressed polynucleotide
 CC (I) encoding a polypeptide (II), ABP67991-ABP67996) comprising: (i) any of
 CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
 CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
 CC sequences that hybridize to (i), under moderately stringent conditions;
 CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
 CC degenerate variants of (i). The compositions and methods of the present
 CC invention are useful for the diagnosis, prevention and/or treatment of
 CC cancer, particularly colon cancer. (i) can be used in gene therapy and
 CC (ii) and (iii) are useful in pharmaceutical compositions such as vaccines.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 CC
 SQ Sequence 505 BP; 120 A; 125 C; 119 G; 141 T; 0 other;

Query Match 19.7%; Score 403; DB 24; Length 505;
 Best Local Similarity 100.0%; Pred. No. 2.6e-189;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1301 GGCTTGGTCCACCTGCTGTCAGCTCAGTCACTTCTCTGAATCTTCTGAGCCTTGC 1360
 DB 1 GGCTTGGTCCACCTGCTGTCAGCTCAGTCACTTCTCTGAATCTTCTGAGCCTTGC 60
 QY 1361 CTACATCTCTGAGACCTTAACACACAGTGTGGGATACACACTGAGTGTGTGTC 1420
 DB 61 CTACATCTCTGAGACCTTAACACACAGTGTGGGATACACACTGAGTGTGTGTC 120
 QY 1421 AACAGTGTCCAGACAG 1480
 DB 121 AACAGTGTCCAGACAG 180
 QY 1481 AACCTTATGTGAGACAGACATCCACAGTGAAGAGGATATGCTCTCTCTTCTTCTT 1540
 DB 181 AACCTTATGTGAGACAGACATCCACAGTGAAGAGGATATGCTCTCTCTTCTTCTT 240
 QY 1541 GATCTTCCCTGCTGGGACAGACTTCAAGACTTGTGGCTGGAGGCTTATTAAGCAGAC 1600
 DB 241 GATCTTCCCTGCTGGGACAGACTTCAAGACTTGTGGCTGGAGGCTTATTAAGCAGAC 300
 QY 1601 ACAGTATCAGTGAATGTGATCCATAAAGCTCCCTGTCCACACATCTGCCAATGGGAATG 1660
 DB 301 ACAGTATCAGTGAATGTGATCCATAAAGCTCCCTGTCCACACATCTGCCAATGGGAATG 360
 QY 1661 GATCTTTCACCAAGAGCTCACCAGATTTTCCACAGAGATGC 1703
 DB 361 GATCTTTCACCAAGAGCTCACCAGATTTTCCACAGAGATGC 403

RESULT 8
 ABR54724
 ID ABR54724 standard; cDNA; 517 BP.
 XX ABR54724;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Human colon cancer-associated cDNA, SEQ ID NO 194.
 XX
 KM Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.
 OS Homo sapiens.
 XX
 XX WO200212280-A2.
 PN
 XX 14-FEB-2002.
 PD

XX 30-JUL-2001; 2001WO-0523826.
 PF
 XX
 PR 03-AUG-2000; 2000US-223265P.
 PR 02-OCT-2000; 2000US-237406P.
 PR 20-MAR-2001; 2001US-277455P.
 PR 03-JUL-2001; 2001US-302702P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Pyle RA, Xu J, Secretist H;
 XX
 DR WPI; 2002-257462/30.
 XX
 PT Novel polynucleotide encoding colon tumour polypeptides, useful as
 XX vaccines for treating colon cancers
 XX
 PS Claim 1; Page 206; 425bp; English.

XX The invention relates to isolated polynucleotides (I) encoding colon
 CC tumour polypeptides (II). (I) is useful for stimulating an immune
 CC response in a patient and treating colon cancer in a patient.
 CC Oligonucleotides derived from (I) are useful for determining the presence
 CC of cancer in a patient. (I) and (II) are useful in pharmaceutical
 CC compositions, e.g. vaccines, and other compositions for the diagnosis
 CC and treatment of colon cancer. A composition comprising a first component
 CC selected from physiologically acceptable carriers and immunostimulants,
 CC and an antigen-presenting cell expressing (II) is useful for inhibiting
 CC development of cancer in a patient. (I) is useful in the design and
 CC preparation of ribozyme molecules for inhibiting expression of tumour
 CC polypeptides and (I). ABR5431-ABR5464 represent human colon cancer cDNA
 CC sequences of the invention.
 CC
 SQ Sequence 517 BP; 121 A; 130 C; 124 G; 142 T; 0 other;

Query Match 19.7%; Score 403; DB 24; Length 517;
 Best Local Similarity 100.0%; Pred. No. 2.6e-189;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1301 GGCTTGGTCCACCTGCTGTCAGCTCAGTCACTTCTCTGAATCTTCTGAGCCTTGC 1360
 DB 13 GGCTTGGTCCACCTGCTGTCAGCTCAGTCACTTCTCTGAATCTTCTGAGCCTTGC 72
 QY 1361 CTACATCTCTGAGACCTTAACACACAGTGTGGGATACACACTGAGTGTGTGTC 1420
 DB 73 CTACATCTCTGAGACCTTAACACACAGTGTGGGATACACACTGAGTGTGTGTC 132
 QY 1421 ACAGTGTCCAGACAG 1480
 DB 133 ACAGTGTCCAGACAG 192
 QY 1481 AACCTTATGTGAGACAGACATCCACAGTGAAGAGGATATGCTCTCTCTTCTTCTT 1540
 DB 193 AACCTTATGTGAGACAGACATCCACAGTGAAGAGGATATGCTCTCTCTTCTTCTT 252
 QY 1541 GATCTTCCCTGCTGGGACAGACTTCAAGACTTGTGGCTGGAGGCTTATTAAGCAGAC 1600
 DB 253 GATCTTCCCTGCTGGGACAGACTTCAAGACTTGTGGCTGGAGGCTTATTAAGCAGAC 312
 QY 1601 ACAGTATCAGTGAATGTGATCCATAAAGCTCCCTGTCCACACATCTGCCAATGGGAATG 1660
 DB 313 ACAGTATCAGTGAATGTGATCCATAAAGCTCCCTGTCCACACATCTGCCAATGGGAATG 372
 QY 1661 GATCTTTCACCAAGAGCTCACCAGATTTTCCACAGAGATGC 1703
 DB 373 GATCTTTCACCAAGAGCTCACCAGATTTTCCACAGAGATGC 415

RESULT 9
 AAC76156
 ID AAC76156 standard; cDNA; 2988 BP.
 XX
 AC AAC76156;

XX 08-FEB-2001 (first entry)
XX Human ORFX ORF1711 polynucleotide sequence SEQ ID NO:3421.
DE
XX
XX Human; open reading frame: ORFX; detection: cytostatic; hepatotropic;
XX vulnary; antipariatic; antiparkinsonian; noctropic; neuroprotective;
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX antiviral; antibacterial; antifungal; antirheumatic; antihypertoid;
XX antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antiinflammatory disease; coagulation;
XX thrombosis; contraceptive; ss.
XX
XX Homo sapiens.
XX
XX WO200056473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000MO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach M;
XX WPI; 2000-602362/57.
XX P-PSDB; AAB41947.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 5; Page 2597-2599; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnary;
XX antipariatic; antiparkinsonian; noctropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX antihypertoid; and antinaemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antiinflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 2988 BP; 613 A; 916 C; 842 G; 617 T; 0 other;
Query Match 15.0%; Score 306; DB 21; Length 2988;
Best Local Similarity 100.0%; Pred. No. 3.7e-141;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 452 TTGTATGCTACATGAGAACCTGGTCCCGGAGACAGTCCAGCTTTTCAGTGGAGAAC 511
DB 1480 TTGTATGCTACATGAGAACCTGGTCCCGGAGACAGTCCAGCTTTTCAGTGGAGAAC 1539
OY 512 AGCCGGGCCCCGTGTCTGTGACCTGCTGTGACATATCCACAGATGAATCATCCCC 571
DB 1540 AGCCGGGCCCCGTGTCTGTGACCTGCTGTGACATATCCACAGATGAATCATCCCC 1599
OY 572 CGGGCTACATGAGAGCTCTGTGACATATCCACAGATGAATCATCCCC 631
DB 1600 CGGGCTACATGAGAGCTCTGTGACATATCCACAGATGAATCATCCCC 1659
OY 632 CGCTCTACAGCAGCAGTGTGCTCAAGAGAGTGGCTTTCTCAACCTGACAGTCCCTAC 691
DB 1660 CGCTCTACAGCAGCAGTGTGCTCAAGAGAGTGGCTTTCTCAACCTGACAGTCCCTAC 1719
OY 692 CGCTCTGTAAGAACCCCTCTCAACCTGATATGCTGACCTGCTCGGAGACCCCGG 751
DB 1720 CGCTCTGTAAGAACCCCTCTCAACCTGATATGCTGACCTGCTCGGAGACCCCGG 1779
OY 752 GCCGTG 757
DB 1780 GCCGTG 1785
RESULT 10
ABR54794
ID ABR54794 standard; cDNA; 389 BP.
XX
XX ABR54794;
XX
XX 18-JUN-2002 (first entry)
XX
XX Human colon cancer-associated cDNA, SEQ ID NO 264.
XX
XX Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200212280-A2.
XX
XX 14-FEB-2002.
XX
XX 30-JUL-2001; 2001MO-US23826.
XX
XX 03-AUG-2000; 2000US-223265P.
XX 02-OCT-2000; 2000US-237406P.
XX 20-MAR-2001; 2001US-277495P.
XX 03-JUL-2001; 2001US-302702P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Pyle RA, Xu J, Secrist H;
XX WPI; 2002-257462/30.
XX
XX Novel polynucleotide encoding colon tumour polypeptides, useful as
XX vaccines for treating colon cancers -
XX
XX Claim 1; Page 225; 425pp; English.
XX
XX The invention relates to isolated polynucleotides (I) encoding colon
XX tumour polypeptides (II). (I) is useful for stimulating an immune
XX response in a patient and treating colon cancer in a patient.
XX Oligonucleotides derived from (I) are useful for determining the presence
XX of cancer in a patient. (I) and (II) are useful in pharmaceutical
XX compositions, e.g. vaccines, and other compositions for the diagnosis
XX and treatment of colon cancer. A composition comprising a first component
XX selected from physiologically acceptable carriers and immunostimulants,
XX and an antigen-presenting cell expressing (II) is useful for inhibiting
XX development of cancer in a patient. (I) is useful in the design and
XX preparation of ribozyme molecules for inhibiting expression of tumour

CC polypeptides and (I). ABR54531-ABR55464 represent human colon cancer cDNA
 CC sequences of the invention.
 XX
 SQ Sequence 389 BP; 121 A; 83 C; 76 G; 109 T; 0 other;

Query Match 5.6%; Score 114; DB 24; Length 389;
 Best Local Similarity 100.0%; Pred. No. 6.9e-46;
 Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1930 TCCCATCTGTATTAATAATCCCAATTAAGGTCCTTTAGATGTCCTTTTATG 1989
 DB 266 TCCCATCTGTATTAATAATCCCAATTAAGGTCCTTTAGATGTCCTTTTATG 325

OY 1990 CTTCTTAATTATTAAGCAATAATGTCATTTTATGGATCTTAATAAAAAA 2043
 DB 326 CTTCTTAATTATTAAGCAATAATGTCATTTTATGGATCTTAATAAAAAA 379

RESULT 11

AAK92588/C
 ID AAK92588 standard; cDNA; 548 BP.

AC AAK92588;

DT 06-NOV-2001 (first entry)

DE Human cDNA 3'-end sequence, SEQ ID NO: 1048.

KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX Homo sapiens.

PN EPI130094-A2.

PD 05-SEP-2001.

PF 07-JUL-2000; 2000EP-0114089.

PR 08-JUL-1999; 99JP-0194486.

PR 11-JAN-2000; 2000JP-0118774.

PR 02-MAY-2000; 2000JP-0183765.

XX (HELI-) HELIX RES INST.

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX WPI; 2001-524255/58.

DR 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -

PS Claim 3; SEQ ID NO 1048; 1380PP + sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been

CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length

CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full

CC length enriched cDNA libraries that were synthesized by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily

CC without any special methods. The present sequence is the nucleotide
 CC sequence of the 3'-end of a cDNA provided in the invention.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 548 BP; 160 A; 109 C; 129 G; 148 T; 2 other;

Query Match 5.1%; Score 105; DB 22; Length 548;
 Best Local Similarity 100.0%; Pred. No. 2e-41;
 Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1930 TCCCATCTGTATTAATAATCCCAATTAAGGTCCTTTAGATGTCCTTTTATG 1989
 DB 108 TCCCATCTGTATTAATAATCCCAATTAAGGTCCTTTAGATGTCCTTTTATG 49

OY 1990 CTTCTTAATTATTAAGCAATAATGTCATTTTATGGATCTTAATAAAAAA 2034
 DB 48 CTTCTTAATTATTAAGCAATAATGTCATTTTATGGATCTTAATAAAAAA 4

RESULT 12

AAAD24670
 ID AAAD24670 standard; cDNA; 1647 BP.

AC AAAD24670;

DT 12-MAR-2002 (first entry)

DE Human drug metabolizing enzyme (DME)-5 cDNA.

XX Human; drug metabolizing enzyme; gene therapy; autoimmune disorder;

KW Inflammatory disorder; acquired immune deficiency syndrome; infection;

KW AIDS; adult respiratory distress syndrome; cell proliferative disorder;

KW allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;

KW cancer; endocrine disorder; hypothalamus disorder; pituitary disorder;

KW gastrointestinal disorder; metabolic disorder; developmental disorder;

KW liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;

KW goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;

XX DME-5; ss.

XX Homo sapiens.

PN Key

PD CDS

PF sig_peptide

PR mat_peptide

PR sig_peptide

PR mat_peptide

PR sig_peptide

PR mat_peptide

PR sig_peptide

PR mat_peptide

PR sig_peptide

PR mat_peptide

PR sig_peptide

PR mat_peptide

PR sig_peptide

PR mat_peptide

PR sig_peptide

PR mat_peptide

PR sig_peptide

PR mat_peptide

PR sig_peptide

PS Claim 5; Page 139; 143pp; English.

XX The invention relates to human drug metabolizing enzymes referred as

CC DME and nucleic acid molecules encoding such enzymes. Polynucleotides

CC of the invention are useful for assessing toxicity of test compounds

CC and in gene therapy. Sequences of the invention are useful in the

CC diagnosis, prevention and treatment of autoimmune/inflammatory

CC disorders such as acquired immune deficiency syndrome (AIDS), adult

CC respiratory distress syndrome, allergies, anaemia, atherosclerosis,

CC asthma, autoimmune haemolytic anaemia, contact dermatitis, Crohn's

CC disease, glomerulonephritis, Goodpasture's syndrome, gout, Grave's

CC disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple

CC sclerosis, osteoarthritis, osteoporosis, psoriasis, systemic lupus

CC erythematosus, rheumatoid arthritis, scleroderma, ulcerative colitis,

CC uveitis, viral, bacterial, fungal, parasitic, protozoal, helminthic

CC infections; cell proliferative disorders such as actinic keratosis,

CC arteriosclerosis, atherosclerosis, Duchenne and Becker dystrophy,

CC hepatitis, Cushing's syndrome, cancers, myelodysplastic syndrome,

CC epilepsy; endocrine disorders such as disorders of the hypothalamus

CC and pituitary resulting from lesions such as primary brain tumours,

CC adenomas, infarction associated with pregnancy, aneurysms, vascular

CC malformations; eye disorders such as conjunctivitis, iritis, retinitis,

CC glaucoma, pigmentosa; metabolic disorders such as Addison's disease,

CC cystic fibrosis, diabetes, goitre, glycogen storage diseases,

CC hypercholesterolaemia, hyperthyroidism, hypoglycaemia, lipid myopathies,

CC Menkes syndrome, mannosidosis, obesity; gastrointestinal disorders such

CC as dysphagia, gastric carcinoma, anorexia, nausea, gastroenteritis,

CC hyperbilirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's

CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental

CC disorders. The present sequence is human DME-5 protein cDNA.

XX

SQ Sequence 1647 BP; 287 A; 560 C; 501 G; 299 T; 0 other;

Query Match 2.9%; Score 59; DB 24; Length 1647;

Best Local Similarity 100.0%; Pred. No. 1.4e-18;

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGCGCTCCTACAGCAGCGTGTGCTCAAGAGAGGCGCTTCTTCAACCTGCAG 682

DB 488 AGGCTGCGCGCTCCTACAGCAGCGTGTGCTCAAGAGAGGCGCTTCTTCAACCTGCAG 546

RESULT 13

AA02700 standard; cDNA; 1694 BP.

XX

AC AA02700;

XX

DT 02-MAY-2001 (first entry)

XX

DE Human glycosyl sulfotransferase-4beta (GST-4beta) cDNA.

XX

KW Human: glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive;

KW therapy; selectin binding inhibitor; gene therapy; inflammation;

KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;

KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;

KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenailitis;

KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;

KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;

KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;

KW asthma; hypersensitivity; rheumatic fever; tissue rejection;

KW chromosome 16q23.1; ss.

XX

OS Homo sapiens.

XX

PH Key Location/Qualifiers

FT 5'UTR 9..188

FT /tag- a

FT CDS 189..1376

FT /tag- b

FT /product- "Human glycosyl sulfotransferase-4beta

FT (GST-4beta)"

FT /note- "CDS is referred as SEQ ID NO:21 in brief

FT description of the figures (page no: 4)"

FT 3'UTR 1377..1694

FT /tag- c

XX

XX W0200106015-A1.

XX

PD 25-JAN-2001.

XX

PF 19-JUL-2000; 2000WO-0519741.

XX

PR 20-JUL-1999; 9905-0144694.

XX

PR 13-JUL-2000; 2000US-0593828.

XX

PA (REGC) UNIV CALIFORNIA.

XX

PI Rosen SD, Lee JK, Hemmerich S;

XX

XX WPI; 2001-138471/14.

DR

DR P-PSDB; AAY72640.

XX

PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for

PT diagnostic and therapeutic agent screening applications

XX

PS Claim 6; Fig 4A; 128pp; English.

XX

CC The present sequence is human glycosyl sulfotransferase-4beta (GST-4

CC beta) cDNA. GST-4 gene is found on chromosome 16q23.1.

CC GST is a type 2 membrane protein useful for inhibiting a binding event

CC between a selectin and a selectin ligand, which comprises contacting the

CC selectin with a non-sulphated selectin ligand, GST and a small molecular

CC agent that inhibits the sulphation activity of GST. GST is also useful

CC in inhibiting a selectin mediated binding event. GST is useful in gene

CC therapy to treat disorders such as acute or chronic inflammation,

CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis

CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,

CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's

CC disease, Grave's disease, adenailitis, hypoparathyroidism, periculous

CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,

CC dermatitis, myocarditis, regional enteritis, adult respiratory distress

CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,

CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection

CC during transplantation.

XX

SQ Sequence 1694 BP; 259 A; 586 C; 514 G; 335 T; 0 other;

Query Match 2.9%; Score 59; DB 22; Length 1694;

Best Local Similarity 100.0%; Pred. No. 1.4e-18;

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGCGCTCCTACAGCAGCGTGTGCTCAAGAGAGGCGCTTCTTCAACCTGCAG 682

DB 676 AGGCTGCGCGCTCCTACAGCAGCGTGTGCTCAAGAGAGGCGCTTCTTCAACCTGCAG 734

RESULT 14

AA02699 standard; cDNA; 2044 BP.

XX

AC AA02699;

XX

DT 02-MAY-2001 (first entry)

XX

DE Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #2.

XX

KW Human: glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;

KW therapy; selectin binding inhibitor; gene therapy; inflammation;

KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;

KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;

KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenailitis;

KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;

KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;

KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;

KW asthma; hypersensitivity; rheumatic fever; tissue rejection;

XX	chromosome 16q23.1; ss.	
OS	Homo sapiens.	
XX		
XX	Key	Location/Qualifiers
XX	Key	218..1390
XX	CDS	/*tag= a
XX		/product= "Human glycosyl sulfotransferase-4alpha
XX		(GST-4alpha)"
XX		/note= "CDS is specifically claimed as SEQ ID NO: 4
XX		in claim 6 (page no: 41) of the specification"
XX		
XX	MO200106015-A1.	
XX		
XX	25-JAN-2001.	
XX		
XX	19-JUL-2000; 2000WO-US19741.	
XX		
XX	20-JUL-1999; 99US-0144694.	
XX		
XX	13-JUL-2000; 2000US-0593828.	
XX		
XX	(REGC) UNIV CALIFORNIA.	
XX		
XX	Rosen SD, Lee JK, Hemmerich S;	
XX		
XX	WPI: 2001-138471/14.	
XX		
XX	P-PSDB: NAY72639.	
XX		
XX	Claim 6; Fig 1; 128pp; English.	
XX		
XX	The present sequence is human glycosyl sulfotransferase-4alpha (GST-4	
XX	alpha) cDNA. GST-4 gene is found on chromosome 16q23.1.	
XX	GST is a type 2 membrane protein useful for inhibiting a binding event	
XX	between a selectin and a selectin ligand, which comprises contacting the	
XX	selectin with a non-sulphated selectin ligand, GST and a small molecular	
XX	agent that inhibits the sulphation activity of GST. GST is also useful	
XX	in inhibiting a selectin mediated binding event. GST is useful in gene	
XX	therapy to treat disorders such as acute or chronic inflammation,	
XX	systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis	
XX	nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,	
XX	glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's	
XX	disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious	
XX	anemia, demyelinating diseases, cirrhosis, ulcerative colitis,	
XX	dermatitis, myocarditis, regional enteritis, adult respiratory distress	
XX	syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,	
XX	bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection	
XX	during transplantation.	
XX	Note: The present sequence is also shown in sequence listing (page	
XX	no: 56) but lacks four nucleotides at its 3' end.	
XX		
XX	Sequence 2044 BP; 422 A; 645 C; 570 G; 407 T; 0 other;	
XX		
XX	Query Match	2.9%; Score 59; DB 22; Length 2044;
XX	Best Local Similarity	100.0%; Pred. No. 1.4e-18;
XX	Matches	59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX		
XX	624 AGGCGTCGGCTCCTACAGCCACGCGTGCCTCAGAGAGTCCGCTTCTTCAACCTGCAG	682
XX		
XX	708 AGGCGTCGGCTCCTACAGCCACGCGTGCCTCAGAGAGTCCGCTTCTTCAACCTGCAG	766
XX		
XX	RESULT 15	
XX	AAD02698	
XX	AAD02698 standard; cDNA; 2170 BP.	
XX		
XX	AAD02698;	
XX		
XX	02-MAY-2001 (first entry)	
XX		
XX	Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #1.	

XX		Human: glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;
KW		therapy; selectin binding inhibitor; gene therapy; inflammation;
KM		systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW		polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KM		glomerulonephritis; myasthenia gravis; Sjogren's syndrome; anemia;
KV		Hashimoto's disease; Grave's disease; hypoparathyroidism; pernicious
KW		demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KM		myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW		asthma; hypersensitivity; rheumatic fever; tissue rejection;
XX		chromosome 16q23.1; ss.
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	5'UTR	9..343
FT	/tag- a	
FT	CDS	344..1516
FT	/tag- b	
FT	/product= "Human glycosyl sulfotransferase-4alpha	
FT	(GST-4alpha)"	
FT	/note= "CDS is specifically claimed as SEQ ID NO: 4	
FT	in claim 6 (page no: 41) of the specification"	
FT	3'UTR	1517..2134
FT	/tag- c	
XX		
PN	WO200106015-A1.	
XX		
PD	25-JAN-2001.	
XX		
PE	19-JUL-2000; 2000WO-US19741.	
XX		
PR	20-JUL-1999; 99US-0144694.	
PR	13-JUL-2000; 2000US-05393828.	
PA	(REGC) UNIV CALIFORNIA.	
PI	Rosen SD, Lee JK, Hemmerich S;	
DR	WPI; 2001-138471/14.	
DR	P-PSDB; AAY72639.	
PT	New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for	
PT	diagnostic and therapeutic agent screening applications -	
XX		
PS	Claim 6; Page 62; 128pp; English.	
XX		
CC	The present sequence is human glycosyl sulfotransferase-4alpha (GST-4	
CC	alpha) cDNA. GST-4 gene is found on chromosome 16q23.1.	
CC	GST is a type 2 membrane protein useful for inhibiting a binding event	
CC	between a selectin and a selectin ligand, which comprises contacting the	
CC	selectin with a non-sulfated selectin ligand, GST and a small molecular	
CC	agent that inhibits the sulphation activity of GST. GST is also useful	
CC	in inhibiting a selectin mediated binding event. GST is useful in gene	
CC	therapy to treat disorders such as acute or chronic inflammation,	
CC	systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis	
CC	nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,	
CC	glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's	
CC	disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious	
CC	anemia, demyelinating diseases, cirrhosis, ulcerative colitis,	
CC	dermatitis, myocarditis, regional enteritis, adult respiratory distress	
CC	syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,	
CC	bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection	
CC	during transplantation.	
XX		
SQ	Sequence 2170 BP; 449 A; 682 C; 612 G; 427 T; 0 other:	
Query Match	2.9%; Score 59; DB 22; Length 2170;	
Best Local Similarity	100.0%; Pred. No. 1,4e-18;	
Matches	59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	

624 AGGCGTGGCTCTACAGCCAGGTGTGTCATCAGAGAGTGCGCTTCAACTGCAG 682

Fri Aug 15 08:09:29 2003

us-09-645-078-1_1.rng

Page 13

Db 834 AGGCTGCGCCTCTACAGCCAGTGTGTCTCAAGAGAGTGTGGCTTTTCACCTGCAG 892

Search completed: August 15, 2003, 01:05:21
Job time : 546 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2003, 00:44:22 ; Search time 3887 Seconds
(without alignments)
12774.388 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043
Sequence: 1 gaattccattgttgtgtgta.....tgggacccctaaaaaaa 2043

Scoring table: OLIGO_NMC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45454738

Minimum DB seq length: 50

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estbda:*
2: em_estbhum:*
3: em_estlnu:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_lnv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	685	33.5	954	BI823850 603039012
2	579	28.3	583	BK490456 Homo sapi
3	534	26.1	668	AL709927 DKFZP6860
4	479	23.4	571	BM129080 1F17C04.Y

Result No.	Score	Query Length	ID	Description
1	685	33.5	954	BI823850 603039012
2	579	28.3	583	BK490456 Homo sapi
3	534	26.1	668	AL709927 DKFZP6860
4	479	23.4	571	BM129080 1F17C04.Y

ALIGNMENTS

RESULT 1
BI823850
LOCUS
DEFINITION
603039012P1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5179826 5', mRNA sequence.
ACCESSION
BI823850
VERSION
BI823850.1 GI:15935400
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 954)
NIH-MGC http://mgi.nci.nih.gov/
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
CONTACT
Robert Strausberg, Ph.D.
Email: gspbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Plate: L1AM1148 row: h column: 03
High quality sequence stop: 856.

FEATURES High quality sequence stop: 434.
Location/Qualifiers
1. 571

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:567655"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="Melton Normalized Human Islet 4 M-HIS 1"

/note="Organ: Pancreas; Vector: pSPORT1; Site: 1; Not 1; Site: 2: Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation: average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 109 a 169 c 143 g 150 t
ORIGIN

Query Match 23.4%; Score 479; DB 12; Length 571;
Best Local Similarity 100.0%; Pred. No. 1.2e-186;
Matches 479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 CAGGCTTCACACTGACACAAATGCTACTGCTTAAATAATGACCTCTGCTGTTCT 165
DB 73 CAGGCTTCACACTGACACAAATGCTACTGCTTAAATAATGACCTCTGCTGTTCT 132
QY 166 GGTTCACAGATGGGCTGCTGCTTATCTTCCACATGATACGCCACACATCAGCTC 225
DB 133 GGTTCACAGATGGGCTGCTGCTTATCTTCCACATGATACGCCACACATCAGCTC 192
QY 226 CCGTCTATGAAGGACACAGCCGAGGAGCATGACGTGCTGCTTCCGCGGCGTC 285
DB 193 CCGTCTATGAAGGACACAGCCGAGGAGCATGACGTGCTGCTTCCGCGGCGTC 252
QY 286 TGGCTCTTCTTTTGTGGGACAGCTTTTGGGACAGCCACAGATGTTTCTACCTGATGA 345
DB 253 TGGCTCTTCTTTTGTGGGACAGCTTTTGGGACAGCCACAGATGTTTCTACCTGATGA 312
QY 346 GCCCGCTGACAGTGTGATGACCTTCAAGCAGACAGCCGCTGATGCTGACATGGC 405
DB 313 GCCCGCTGACAGTGTGATGACCTTCAAGCAGACAGCCGCTGATGCTGACATGGC 372
QY 406 TGTGGGATCTGATACGGGCGCTTCTTGTGCGACATGAGCGCTTTGATGCTACAT 465
DB 373 TGTGGGATCTGATACGGGCGCTTCTTGTGCGACATGAGCGCTTTGATGCTACAT 432
QY 466 GGAACCTGTCCTCCGAGACAGTCAAGCTCTTTCAAGTGGGAGACAGCCGCGCTGTG 525
DB 433 GGAACCTGTCCTCCGAGACAGTCAAGCTCTTTCAAGTGGGAGACAGCCGCGCTGTG 492
QY 526 TTCTGACCTGCTGATCATCCCAAAATGAAATCAATCCCGGGGCTCAGTCA 584
DB 493 TTCTGACCTGCTGATCATCCCAAAATGAAATCAATCCCGGGGCTCAGTCA 551

RESULT 5
LOCUS BM969292 593 bp mRNA linear EST 20-FEB-2003
DEFINITION UI-CF-ENO-acp-1-21-0-UI.s1 UI-CF-ENO Homo sapiens cDNA clone
ACCESSION BM969292
VERSION BM969292.1 GI:19586879

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (Bases 1 to 593)
Bonaldo,M.F., Lennon,G. and Soares,M.B.

JOURNAL Normalization and subtraction: two approaches to facilitate gene
MEDLINE Genome Res. 6 (9), 791-806 (1996)
PUBMED 97044477
8889548

COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171

Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com) or from Open Biosystems (www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA=yes.

FEATURES Location/Qualifiers
1. 593

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-ENO-acp-1-21-0-UI"
/tissue_type="Primary Lung Cystic Fibrosis Epithelial Cells"

/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-ENO"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site: 1: EcoR I; Site: 2: Not I; UI-CF-ENO is a cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGCT.
TAG LIB=UI-CF-ENO
TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS 6hr to LPS 24h
TAG_SBO=CTGCTCAGCT"

BASE COUNT 157 a 127 c 137 g 172 t
ORIGIN

Query Match 15.8%; Score 322; DB 12; Length 593;
Best Local Similarity 100.0%; Pred. No. 6.6e-122;
Matches 322; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1382 ACATGCTGTGGTATCACACTGAGTGTGTCACAGTGCCTCAGACAGAAGA 1441
DB 593 ACATGCTGTGGTATCACACTGAGTGTGTCACAGTGCCTCAGACAGAAGA 534
QY 1442 CTTTGTGTTCATGCTTGTGTAGAAAACAGACTGGGAACTTATGTGACACACAT 1501
DB 533 CTTTGTGTTCATGCTTGTGTAGAAAACAGACTGGGAACTTATGTGACACACAT 474
QY 1502 CCCACCACTGAAACAGAGGATATGCTCTTCTTTCTTGTGATCTTCTGCGCAGAC 1561

|||||
 Db 473 CCCACGATGAAACAGGGTATGCTCTTCTTTCTTGTATCTTCTGCTGCGGACAGAC 414
 OY 1562 TTTCAGACATTTGGGCTGAGGCGCTATTAAGACACACACAGATATAGGATATATC 1621
 Db 413 TTTCAGACATTTGGGCTGAGGCGCTATTAAGACACACAGATATAGGATATATC 354
 OY 1622 CATTAACCTCCCTGTCACATCTTGGCCAAATGGGATGATCTTTCACCAAGAAGCTCA 1681
 Db 353 CATTAACCTCCCTGTCACATCTTGGCCAAATGGGATGATCTTTCACCAAGAAGCTCA 294
 OY 1682 CCAGCATTTTCCACAGAGATGC 1703
 Db 293 CCAGCATTTTCCACAGAGATGC 272

RESULT 6
 BF878439
 LOCUS BE878439 680 bp mRNA linear EST 17-JAN-2001
 DEFINITION MR0-ET0109-191100-002-h06 ET0109 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF878439
 VERSION BF878439.1 GI:12268569
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 680)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordu, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 PUBMED 10737800
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the PAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0<2=MR0-ET0109-191100-002-h06<3=2000-11-19<4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 17
 High quality sequence stop: 641.
 Location/Qualifiers
 1. 680
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="adult"
 /clone_1lb="ET0109"
 /note="Organ: lung_tumor; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent Application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 168 a 164 c 163 g 184 t 1 others
 ORIGIN
 Query Match 14.2%; Score 290; DB 10; Length 680;
 Best Local Similarity 100.0%; Pred. No. 9.9e-109;

Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1414 TGTGTCCACAGCTGCTCAGAGAGAGACTTTTGTGTCATCGCTGTGTAGAAACAG 1473
 Db 106 TGTGTCCACAGCTGCTCAGAGAGAGACTTTTGTGTCATCGCTGTGTAGAAACAG 165
 OY 1474 ACTGGGAACTTATGTGACAGACATCCACCAGTAAACAGGATATGCTCTTTC 1533
 Db 166 ACTGGGAACTTATGTGACAGACATCCACCAGTAAACAGGATATGCTCTTTC 225
 OY 1534 TTTTCTGATCTTCTGCTGCGGAGACTTCAGAGACTTGTGCGCTGAGGCGCTATTAA 1593
 Db 226 TTTTCTGATCTTCTGCTGCGGAGACTTCAGAGACTTGTGCGCTGAGGCGCTATTAA 285
 OY 1594 GCACGACACAGTATCATGTGAATGATCCATAAACCTCCCTGTCACATCTTGCCCAATG 1653
 Db 286 GCACGACACAGTATCATGTGAATGATCCATAAACCTCCCTGTCACATCTTGCCCAATG 345
 OY 1654 GGGATGATCTTTCACCAAGAGCTCAGACAGATTTTCCACAGAGATGC 1703
 Db 346 GGGATGATCTTTCACCAAGAGCTCAGACAGATTTTCCACAGAGATGC 395

RESULT 7
 BU684395/c
 LOCUS BU684395 669 bp mRNA linear EST 07-OCT-2002
 DEFINITION UI-CF-ENO-aco-f-08-0-UI-s1 UI-CF-ENO Homo sapiens cDNA clone
 UI-CF-ENO-aco-f-08-0-UI 3', mRNA sequence.
 ACCESSION BU684395
 VERSION BU684395.1 GI:23537302
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 669)
 Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 PUBMED 8889548
 COMMENT Contact: McCray, PB
 McCray Lab
 University of Iowa
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
 Tel: 319 356 4866
 Fax: 319 356 7171
 Email: paul.mccray@uiowa.edu
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa
 DNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com) or from Open Biosystems (www.openbiosystems.com).
 Seq primer: M13 FORWARD
 Polya=Yes.
 Location/Qualifiers
 1. 669
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_1lb="UI-CF-ENO"
 /note="Organ: lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-CF-ENO is a cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells."
 FEATURES
 source

JOURNAL
COMMENT

Unpublished
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biochem.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brownjefas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
High quality sequence stop: 425.

FEATURES

SOURCE

1..553

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5676655"

/sex="Both"

/tissue_type="Islets of Langerhans"

/dev_stage="Adult"

/lab_host="DH10B"

/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
Site_2: Sal 1; Starting library constructed using
SuperScript Plasmid Library Kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an EcoT of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."

BASE COUNT

163 a 111 c 127 g 152 t

ORIGIN

Query Match 10.7%; Score 218; DB 12; Length 553;
Best Local Similarity 100.0%; Pred. No. 5.3e-79;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1486 TATGTAGCAGACATCCACAGGATGTAAGGCTTCTTTCTTTCTGATCT 1545
DB 553 TATGTAGCAGACATCCACAGGATGTAAGGCTTCTTTCTTTCTGATCT 494
QY 1546 TCCGTCTGGGACACTTCAGAGACTTGGGCTTGAGGCTTAAAGCAGCAGACT 1605
DB 493 TCCGTCTGGGACACTTCAGAGACTTGGGCTTGAGGCTTAAAGCAGCAGACT 434
QY 1606 ATCAGTGAATTGATTCATAAAGCTCCCTGTCCACATCTGGCCAAATGGGAAGGATCT 1665
DB 433 ATCAGTGAATTGATTCATAAAGCTCCCTGTCCACATCTGGCCAAATGGGAAGGATCT 374
QY 1666 TTCACCAAGAGCTCACCAGCATTTCACAGAGATGC 1703
DB 373 TTCACCAAGAGCTCACCAGCATTTCACAGAGATGC 336

RESULT 10

AM002418/c

LOCUS

DEFINITION AM002418 362 bp mRNA linear EST 27-OCT-1999

ACCESSION w61903.x1 NCI_CGAP_G66 Homo sapiens cDNA clone IMAGE:2524564 3',

VERSION AM002418 mRNA sequence.

KEYWORDS AM002418.1 GI:5849334

SOURCE EST.
Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
TITLE
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL
COMMENT

Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrp/image/image.html
Seq primer: -40UP from gibco.
Location/Qualifiers

FEATURES

SOURCE

1..362

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2524564"

/tissue_type="pooled germ cell tumors"

/lab_host="DH10B"

/clone_id="NCI_CGAP_G66"

/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not 1; Site_2: Eco RI; Plasmid DNA
from the normalized library NCI_CGAP_G64 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneids
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

108 a 69 c 77 g 108 t

ORIGIN

Query Match 5.2%; Score 107; DB 9; Length 362;
Best Local Similarity 100.0%; Pred. No. 3.7e-33;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1930 TGGCATCTGTTAACTAAATATCCCAATAGGTTCTGTTAGAAATGCCCTTTATG 1989
DB 107 TGGCATCTGTTAACTAAATATCCCAATAGGTTCTGTTAGAAATGCCCTTTATG 48
QY 1990 CTCTTAATATTTAGCAGTAATGTTCATTTTATGGATCTTAA 2036
DB 47 CTCTTAATATTTAGCAGTAATGTTCATTTTATGGATCTTAA 1

RESULT 11

AM572510/c

LOCUS

DEFINITION AM572510 417 bp mRNA linear EST 13-MAR-2000

ACCESSION x418911.x2 NCI_CGAP_Ur1 Homo sapiens cDNA clone IMAGE:2751044 3',

VERSION AM572510 mRNA sequence.

KEYWORDS AM572510.1 GI:7237243

SOURCE EST.
Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

TITLE Tumor Gene Index

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Cloning distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml
 Seq primer: -40UP from Gibco
 High quality sequence stop: 416.
 Location/Qualifiers

FEATURES

source

1.417
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2751044"
 /tissue_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"
 /lab_host="DH10B"
 /clone_1lb="NCI_CGAP_0t1"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1: SalI; Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"
 BASE COUNT 140 a 65 c 80 g 132 t
 ORIGIN

Query Match 5.1%; Score 105; DB 9; Length 417;
 Best Local Similarity 100.0%; Pred. No. 2.3e-32;
 Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1930 TCCATCTGTATTAATAAATCCCAATAAGTTCGTTAGAAATGTCCTTTTATG 1989
 DB 266 TCCCATCTGTATTAATAAATCCCAATAAGTTCGTTAGAAATGTCCTTTTATG 207
 QY 1990 CTCTTAATTATAGAGTAATGTCATTTTATGCGATCCAA 2034
 DB 206 CTCTTAATTATAGAGTAATGTCATTTTATGCGATCCAA 162

RESULT 12 358 bp mRNA linear EST 16-OCT-2000
 LOCUS BF056840/c 7k10d12.x1 NCI_CGAP_G6 Homo sapiens CDNA clone IMAGE:3443783 3',
 DEFINITION mRNA sequence.
 ACCESSION BF056840 GI:10810736
 VERSION BF056840
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 358)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapsb-r@mail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Cloning distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: infelimage@llnl.gov
 Seq primer: -40UP from Gibco.
 Location/Qualifiers
 1.358
 /organism="Homo sapiens"
 /mol_type="mRNA"

FEATURES

source

1.358
 /organism="Homo sapiens"
 /mol_type="mRNA"

/db_xref="taxon:9606"
 /clone="IMAGE:3443783"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /clone_1lb="NCI_CGAP_G6"
 /note="Vector: p7773D-Pac (Pharmacia) with a modified polylinker; Site: 1: Not I; Site: 2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_G4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 107 a 68 c 82 g 101 t
 ORIGIN

Query Match 4.9%; Score 101; DB 10; Length 358;
 Best Local Similarity 100.0%; Pred. No. 1.1e-30;
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1932 CCATCGTAAATACAAATTCACAAATAGGTCGTTAGAAATGTCCTTTATGCT 1991
 DB 101 CCATCGTAAATACAAATTCACAAATAGGTCGTTAGAAATGTCCTTTATGCT 42
 QY 1992 TCTTAATTATAGCAGTAATGTCATTTTATGCGATCCT 2032
 DB 41 TCTTAATTATAGCAGTAATGTCATTTTATGCGATCCT 1

RESULT 13 419 bp mRNA linear EST 13-MAR-2000
 LOCUS AW572390/c xU09a06.x2 NCI_CGAP_C014 Homo sapiens CDNA clone IMAGE:2799634 3',
 DEFINITION mRNA sequence.
 ACCESSION AW572390 GI:7237123
 VERSION AW572390
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 419)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapsb-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Cloning distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml
 Seq primer: -40UP from Gibco
 High quality sequence stop: 343.
 Location/Qualifiers
 1.419
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2799634"
 /tissue_type="moderately-differentiated adenocarcinoma"
 /lab_host="DH10B"
 /clone_1lb="NCI_CGAP_C014"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site: 1: SalI; Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.7 kb. Life Technologies catalog #: 11531-019"

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapsb-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Cloning distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml
 Seq primer: -40UP from Gibco
 High quality sequence stop: 343.
 Location/Qualifiers
 1.419
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2799634"
 /tissue_type="moderately-differentiated adenocarcinoma"
 /lab_host="DH10B"
 /clone_1lb="NCI_CGAP_C014"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site: 1: SalI; Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.7 kb. Life Technologies catalog #: 11531-019"

FEATURES

source

1.419
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2799634"
 /tissue_type="moderately-differentiated adenocarcinoma"
 /lab_host="DH10B"
 /clone_1lb="NCI_CGAP_C014"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site: 1: SalI; Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.7 kb. Life Technologies catalog #: 11531-019"
 BASE COUNT 122 a 78 c 89 g 130 t

ORIGIN

Query Match 3.6%; Score 74; DB 9; Length 419;
 Best Local Similarity 100.0%; Pred. NO. 1.4e-19;
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1630 TCCCTGTCACATCTTGCACATGGGAGATGATCTTTACCAAGAGCTCACCAGATT 1689
 |||||||
 DB 419 TCCCTGTCACATCTTGCACATGGGAGATGATCTTTACCAAGAGCTCACCAGATT 360
 |||||||

QY 1690 TTCACAGAGATGC 1703
 |||||||

DB 359 TTCACAGAGATGC 346

RESULT 14
 BFI97521/c 525 bp mRNA linear EST 03-NOV-2000
 LOCUS 7684a08.x1 NCI-CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3642903 3'
 DEFINITION similar to TR:Q9UBY3 Q9UBY3 N-ACETYLGLUCOSAMINE
 6-O-SULFOTRANSFERASE. ; mRNA sequence.

ACCESSION BFI97521
 VERSION BFI97521.1 GI:11086670
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 525)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL, send email to:
lnl@image.lnl.gov
 High quality sequence stop: 451.

FEATURES
 source
 Location/Qualifiers
 1..525
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3642903"
 /lab_host="DH10B"
 /clone_lib="NCI-CGAP_Kid11"
 /note="Organ: kidney; Vector: p773D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 Plasmid DNA from the normalized library NCI-CGAP_Kid3 was
 prepared, and ss circles were made in vitro. Following NRP
 purification, this DNA was used as tracer in a subtractive
 hybridization reaction. The driver was PCR-amplified cDNAs
 from a pool of 5,000 clones made from the same library
 (clones 1322376-1323911, 1456007-1456775, and
 1500552-1502855). Subtraction by Bento Soares and M.
 Fatima Bonaldo."

BASE COUNT 90 a 163 c 187 g 85 t

ORIGIN

Query Match 2.9%; Score 59; DB 10; Length 525;
 Best Local Similarity 100.0%; Pred. No. 2e-13;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGCTCTCTACAGCAGCTGCTGCTCAAGAGGCGCTTCTTCAACCTGCAG 682
 |||||||
 DB 260 AGGCTGCGCTCTCTACAGCAGCTGCTGCTCAAGAGGCGCTTCTTCAACCTGCAG 202

RESULT 15
 A1824100/c 620 bp mRNA linear EST 21-DEC-1999
 LOCUS w146c01.x1 NCI-CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2405856 3'
 DEFINITION similar to TR:075667 075667 D7J116.4 ; contains PTR5.b2 PTR5
 repetitive element ; mRNA sequence.

ACCESSION A1824100
 VERSION A1824100.1 GI:5444771
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 620)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL, at:
www-bio.lnl.gov/bdrp/image/image.html
 Insert length: 917 Std Error: 0.00
 Seq primer: -40UP from Glbco
 High quality sequence stop: 490.

FEATURES
 source
 Location/Qualifiers
 1..620
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2405856"
 /tissue="IMAGE:squamous cell carcinoma, poorly
 differentiated (4 pooled tumors, including primary and
 metastatic)"
 /dev_stage="adult"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NCI-CGAP_Lu19"
 /note="Organ: lung; Vector: p773D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 pooled lung tumor tissue, and was then primed with a Not I
 - oligo(dT) primer. Double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 p773 vector. Library went through one round of
 normalization. Library constructed by Bento Soares and M.
 Fatima Bonaldo."

BASE COUNT 106 a 195 c 227 g 89 t 3 others

ORIGIN

Query Match 2.9%; Score 59; DB 9; Length 620;
 Best Local Similarity 100.0%; Pred. No. 1.9e-13;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGCTCTCTACAGCAGCTGCTGCTCAAGAGGCGCTTCTTCAACCTGCAG 682
 |||||||
 DB 314 AGGCTGCGCTCTCTACAGCAGCTGCTGCTCAAGAGGCGCTTCTTCAACCTGCAG 256

Search completed: August 15, 2003, 04:13:33
 Job time : 3888 secs

TT SMTT
EGAD B7A
XMAK
(0000000000)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2003, 00:46:57 ; Search time 141 Seconds
(without alignments)
6395.360 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043

Sequence: 1 gattccatcgtctggtgta.....tggatcctaataaaaaa 2043

Scoring table: OLIGO_NUC

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters:

404290

Minimum DB seq length: 50

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/prodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/prodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/prodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/prodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/prodata/1/lna/PCTUS.COMB.seq:*
6: /cgn2_6/prodata/1/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score-distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1547	75.7	2032	3	US-09-045-284A-1
2	1547	75.7	2032	4	US-09-190-911-2
3	1153	56.4	2065	4	US-09-786-240-26
4	19	0.9	3792	2	US-08-992-334-1
5	19	0.9	3792	2	US-08-302-752-1
6	19	0.9	5234	2	US-08-992-334-2
7	19	0.9	5234	2	US-08-302-752-2
8	19	0.9	6722	2	US-08-992-334-3
9	19	0.9	6722	3	US-08-302-752-3
10	18	0.9	150	4	US-09-439-313-320
11	18	0.9	150	4	US-09-352-616A-320
12	18	0.9	150	4	US-09-232-149A-320
13	18	0.9	201	3	US-09-020-956-78
14	18	0.9	201	3	US-09-030-607-78
15	18	0.9	201	3	US-09-439-313-78
16	18	0.9	201	4	US-09-352-616A-78
17	18	0.9	201	4	US-09-232-149A-78
18	18	0.9	202	3	US-09-030-607-180
19	18	0.9	202	4	US-09-439-313-180
20	18	0.9	202	4	US-09-352-616A-180
21	18	0.9	202	4	US-09-232-149A-180
22	18	0.9	536	3	US-09-020-956-69
23	18	0.9	536	3	US-09-030-607-69
24	18	0.9	536	4	US-09-439-313-69
25	18	0.9	536	4	US-09-352-616A-69
26	18	0.9	536	4	US-09-232-149A-69
27	18	0.9	699	4	US-09-107-532A-1333

28	18	0.9	822	3	US-09-020-956-29	Sequence 29, App1
29	18	0.9	822	3	US-09-030-607-29	Sequence 29, App1
30	18	0.9	822	4	US-09-439-313-29	Sequence 29, App1
31	18	0.9	822	4	US-09-352-616A-29	Sequence 29, App1
32	18	0.9	822	4	US-09-232-149A-29	Sequence 29, App1
33	18	0.9	1197	4	US-09-257-179-26	Sequence 26, App1
34	18	0.9	1368	4	US-09-118-554-61	Sequence 61, App1
35	18	0.9	1368	4	US-09-118-527-61	Sequence 61, App1
36	18	0.9	1368	4	US-09-602-877A-61	Sequence 61, App1
37	18	0.9	1913	4	US-09-589-360B-40	Sequence 40, App1
38	18	0.9	1927	4	US-09-336-536-66	Sequence 66, App1
39	18	0.9	4351	4	US-09-634-238-25	Sequence 25, App1
40	17	0.8	141	1	US-08-036-555B-161	Sequence 161, App
41	17	0.8	141	1	US-08-469-569-161	Sequence 161, App
42	17	0.8	141	1	US-08-249-322A-161	Sequence 161, App
43	17	0.8	141	1	US-08-469-526A-161	Sequence 161, App
44	17	0.8	141	2	US-08-734-591A-161	Sequence 161, App
45	17	0.8	141	2	US-08-469-660-161	Sequence 161, App

ALIGNMENTS

RESULT 1									
US-09-045-284A-1									
Sequence 1, Application US/09045284A									
Patent No. 6265192									
GENERAL INFORMATION:									
APPLICANT: Bistup, Annette									
APPLICANT: Rosen, Steven D.									
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3									
FILE REFERENCE: 6510-107US1									
CURRENT APPLICATION NUMBER: US/09/045, 284A									
CURRENT FILING DATE: 1998-03-20									
NUMBER OF SEQ ID NOS: 9									
SOFTWARE: FastSeq for Windows Version 4.0									
SEQ ID NO 1									
LENGTH: 2032									
TYPE: DNA									
ORGANISM: Homo sapiens									
US-09-045-284A-1									
Query Match									
Best Local Similarity 99.9%; Pred. No. 0;									
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
75.7%; Score 1547; DB 3; Length 2032;									
QY	106	CAAGCTCTCCACTTACGACACATGCTACTGCTTAATAAATGAGCTCTGCTTTCT	165						
DB	160	CAAGCTCTCCACTTACGACACATGCTACTGCTTAATAAATGAGCTCTGCTTTCT	219						
QY	166	GGTTTCCACATGGCCATCTTGGCTATCTTCCATGATGACGACCAACATCAGCTC	225						
DB	220	GGTTTCCACATGGCCATCTTGGCTATCTTCCATGATGACGACCAACATCAGCTC	279						
QY	226	CCGTCTATGAAGCAGCCGAGCCATGACGCTGCTGCTTCTCTGCGCCCTC	285						
DB	280	CCGTCTATGAAGCAGCCGAGCCATGACGCTGCTGCTTCTCTGCGCCCTC	339						
QY	286	TGGCTCTCTTGTGGGGAGCTTTTGGGACACCCAGATTTTCTACCTGATGGA	345						
DB	340	TGGCTCTCTTGTGGGGAGCTTTTGGGACACCCAGATTTTCTACCTGATGGA	399						
QY	346	GCCGCGCTGGCAGCTGTGATGACCTTCAAGCAGCAGCCCTGATGCTGCACATGCC	405						
DB	400	GCCGCGCTGGCAGCTGTGATGACCTTCAAGCAGCAGCCCTGATGCTGCACATGCC	459						
QY	406	TGTCGGGATCTGATGAGGCGCTCTGTGAGCATGAGCCTTTGATGCTTCAT	465						
DB	460	TGTCGGGATCTGATGAGGCGCTCTGTGAGCATGAGCCTTTGATGCTTCAT	519						
QY	466	GGAACCTGTCCTCCGAGACAGTCCACCTCTTTCAGTGGAGAACGCGGCGCTGTG	525						

```

Db      520 GGAACCTGGTCCCGGAGACAGTCCAGCCCTTTTCATGAGGAGAACAGCCGGGCCCTGTG 579
Qy      526 TTCTGCACCTGGCTGTGACATCATCCACAAAGATGAATCAATCCCCGGGCTACTGCAG 585
Db      580 TTCTGCACCTGGCTGTGACATCATCCACAAAGATGAATCAATCCCCGGGCTACTGCAG 639
Qy      586 GCTCTGTGACAGTCAACAGCCCTTTGAGGTGTGAGAAAGCCCTGCCCTCTACAGCA 645
Db      640 GCTCTGTGACAGTCAACAGCCCTTTGAGGTGTGAGAAAGCCCTGCCCTCTACAGCA 699
Qy      646 GGTGGGTCAAGAGGAGGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTGAAA 705
Db      700 GGTGGGTCAAGAGGAGGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTGAAA 759
Qy      706 CCCCTCCCTCAACCTGCATATGCTGACAGCTGTGCGGAGCCCGGGCGGTGTCCGTT 765
Db      760 CCCCTCCCTCAACCTGCATATGCTGACAGCTGTGCGGAGCCCGGGCGGTGTCCGTT 819
Qy      766 CCGAGAACGACAAAGGAGATCTCATGATTGACAGTCCCATTTGTGATGGGCGAGCATGA 825
Db      820 CCGAGAACGACAAAGGAGATCTCATGATTGACAGTCCCATTTGTGATGGGCGAGCATGA 879
Qy      826 GCAAAACCTCAAGAGGAGGAGCAACCCACTATGATGACAGGTGATCTGCCAAAGCA 885
Db      880 GCAAAACCTCAAGAGGAGGAGCAACCCACTATGATGACAGGTGATCTGCCAAAGCA 939
Qy      886 GCTGAGATCTCAAGAGGAGGAGTCTTGGCCCAAGGCGCTGACAGAGCTACTGCT 945
Db      940 GCTGAGATCTCAAGAGGAGGAGTCTTGGCCCAAGGCGCTGACAGAGCTACTGCT 999
Qy      946 TGTGGCTATGAGAGCTGTGCTGAGCCCTGTGGCCGACAGTCCCGAATGTATGAT 1005
Db      1000 TGTGGCTATGAGAGCTGTGCTGAGCCCTGTGGCCGACAGTCCCGAATGTATGAT 1059
Qy      1006 CTTGGGATTTGAAATCTTGGCCCATCTTCTGAGACCTGGGAGCAATACATACCCGAGCA 1065
Db      1060 CTTGGGATTTGAAATCTTGGCCCATCTTCTGAGACCTGGGAGCAATACATACCCGAGCA 1119
Qy      1066 GGGCATGGGTGACACAGCTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGC 1125
Db      1120 GGGCATGGGTGACACAGCTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGC 1179
Qy      1126 TTGGGGCTGTGCTTGGCCCATGAAAGGTTCTGAGTCTGAGAAAGCCCTGTGGCATGC 1185
Db      1180 TTGGGGCTGTGCTTGGCCCATGAAAGGTTCTGAGTCTGAGAAAGCCCTGTGGCATGC 1239
Qy      1186 CATGAATTTGCTGGGTACCGCACAGTCTGATCTGAAACAAAGAGAGAACTGTGCT 1245
Db      1240 CATGAATTTGCTGGGTACCGCACAGTCTGATCTGAAACAAAGAGAGAACTGTGCT 1299
Qy      1246 GATCTCTTCTTACCTGAGCTGTCTCTGAGCAATCCACTAAGAGGTTGAGAGGCTT 1305
Db      1300 GATCTCTTCTTACCTGAGCTGTCTCTGAGCAATCCACTAAGAGGTTGAGAGGCTT 1359
Qy      1306 TGTGTGACCTGTGTGACAGTCAAGTCTTCTGATGCTGTGAGCCCTGTGCTTAA 1365
Db      1360 TGTGTGACCTGTGTGACAGTCAAGTCTTCTGATGCTGTGAGCCCTGTGCTTAA 1419
Qy      1366 TGTGTGACCTTAACTACATGTCTGTGGTATCACACTGATGTGATGTCTTCAACAG 1425
Db      1420 TGTGTGACCTTAACTACATGTCTGTGGTATCACACTGATGTGATGTCTTCAACAG 1479
Qy      1426 TGTCTAAGAGAGAGCTTTTGTCTCAGTCTTGTGTCTAGAAACAGCTGGGAGCT 1485
Db      1480 TGTCTAAGAGAGAGCTTTTGTCTCAGTCTTGTGTCTAGAAACAGCTGGGAGCT 1539
Qy      1486 TATGTGAGAGACATCCACAGTGAAGAGGATTTGCTTCTTCTTCTTGTGATCT 1545
Db      1540 TATGTGAGAGACATCCACAGTGAAGAGGATTTGCTTCTTCTTCTTGTGATCT 1599
Qy      1546 TCTCTTCTGGGAGAGCTTCTAGAGCTTGTGTGCTGTGAGAGCTTATTAAGCAGACAGCT 1605
Db      1600 TCTCTTCTGGGAGAGCTTCTAGAGCTTGTGTGCTGTGAGAGCTTATTAAGCAGACAGCT 1659

```

```

Qy      1606 ATCAGTGAATTGATCCATAAACCTCCCTGTCACATCTTGGCCAAATGGGAGATGATCT 1665
Db      1660 ATCAGTGAATTGATCCATAAACCTCCCTGTCACATCTTGGCCAAATGGGAGATGATCT 1719
Qy      1666 TTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
Db      1720 TTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 1757

RESULT 2
US-09-190-911-2
; Sequence 2, Application US/09190911
; Patent No. 636365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerlich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/09/190,911
; EARLIER FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens
US-09-190-911-2

Query Match      75.7%; Score 1547; DB 4; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      106 CAAGTCTTCCACTTCAGACAAATGCTACTGCTTAAAAAATGAAAGCTCTGCTTTTCT 165
Db      160 CAAGTCTTCCACTTCAGACAAATGCTACTGCTTAAAAAATGAAAGCTCTGCTTTTCT 219
Qy      166 GGTTCGCAAGAGGCGCATCTGGCTATTTTCCAGATGTACAGCACAAATCAGCTC 225
Db      220 GGTTCGCAAGATGCGCATCTGGCTATTTTCCAGATGTACAGCACAAATCAGCTC 279
Qy      226 CCTGTATGAAGGACAGCCGAGGCGATGACAGTGTGTTCTGTTCTTCTGTGGCGCTC 285
Db      280 CCTGTATGAAGGACAGCCGAGGCGATGACAGTGTGTTCTGTTCTTCTGTGGCGCTC 339
Qy      286 TGGCTTTCTTTTGTGGGCGAGCTTTTGGGACAGCCAGATGTTTCTTACTGATGCA 345
Db      340 TGGCTTTCTTTTGTGGGCGAGCTTTTGGGACAGCCAGATGTTTCTTACTGATGCA 399
Qy      346 GCCCGCTGGACAGTGTGATGACCTTCAAGCAGACACCGGCTGTGATGCTGACATGGC 405
Db      400 GCCCGCTGGACAGTGTGATGACCTTCAAGCAGACACCGGCTGTGATGCTGACATGGC 459
Qy      406 TGTGCGGATCTGATACGGGCGCTTCTTGTGCGACATGAGCGTCTTGTGATGCTACAT 465
Db      460 TGTGCGGATCTGATACGGGCGCTTCTTGTGCGACATGAGCGTCTTGTGATGCTACAT 519
Qy      466 GGAACCTGGTCCCCGGAGAGCATGCGAGCTTTTCAAGTGGGAGAACAGCCGGGCGCTGTG 525
Db      520 GGAACCTGGTCCCCGGAGAGCATGCGAGCTTTTCAAGTGGGAGAACAGCCGGGCGCTGTG 579
Qy      526 TTCTGCACCTGGCTGTGACATCATCCACAAAGATGAATCAATCCCCGGGCTACTGCAG 585
Db      580 TTCTGCACCTGGCTGTGACATCATCCACAAAGATGAATCAATCCCCGGGCTACTGCAG 639
Qy      586 GCTCTGTGACAGTCAACAGCCCTTTGAGGTGTGAGAAAGCCCTGCCCTCTACAGCA 645
Db      640 GCTCTGTGACAGTCAACAGCCCTTTGAGGTGTGAGAAAGCCCTGCCCTCTACAGCA 699

```



```

Db      691  ACGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACGTCCTTACCCGCTGCTGAAG 750
QY      705  ACCCTCCCTCAACCTGCATATCTGACACCTGATCCGGAGCCCGGGCGGTTCGTT 764
Db      751  ACCCTCCCTCAACCTGCATATCTGACACCTGATCCGGAGCCCGGGCGGTTCGTT 810
QY      765  CCCGAGAACGCAAAAGGAGATCTCAGATGATGACAGTGCATTTGATGGGAGCATG 824
Db      811  CCCGAGAACGCAAAAGGAGATCTCAGATGATGACAGTGCATTTGATGGGAGCATG 870
QY      825  ACCAAACTCAAGAGAGAGAGCAACCTTACTATGTATGAGAGCATCTGCCAAGCC 884
Db      871  ACCAGAACTCAAGAGAGAGAGCAACCTTACTATGTATGAGAGCATCTGCCAAGCC 930
QY      885  AACTGAGATCTCAAGAGAGAGAGCAACCTTACTATGTATGAGAGCATCTGCCAAGCC 944
Db      931  AACTGAGATCTCAAGAGAGAGAGCAACCTTACTATGTATGAGAGCATCTGCCAAGCC 990
QY      945  TTGTGCGCTATAGAGACCTGGCTCGAGCCCTGTGGCCAGACACTTCCGAAATGATGAA 1004
Db      991  TTGTGCGCTATAGAGACCTGGCTCGAGCCCTGTGGCCAGACACTTCCGAAATGATGAA 1050
QY      1005  TGTGGGATTTGAATTTCTTGGCCCACTTCAAGACCTGGGTGATATACATCACCCGAGGA 1064
Db      1051  TGTGGGATTTGAATTTCTTGGCCCACTTCAAGACCTGGGTGATATACATCACCCGAGGA 1110
QY      1065  AGGCGATGGGTGACCAAGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGG 1124
Db      1111  AGGCGATGGGTGACCAAGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGG 1170
QY      1125  CTTGGCGTGGTCTTGGCCCACTTCAAGACCTGGGTGATATACATCACCCGAGGATG 1184
Db      1171  CTTGGCGTGGTCTTGGCCCACTTCAAGACCTGGGTGATATACATCACCCGAGGATG 1230
QY      1185  CCATGAATTTGCTGGGCTACCCGCAAGTGCAGATCTGACAGAGAGAGAGAGAGAGAG 1244
Db      1231  CCATGAATTTGCTGGGCTACCCGCAAGTGCAGATCTGACAGAGAGAGAGAGAGAGAG 1290
QY      1245  TGGATCTTGTCTGACCTGAGCTGTCGAGCAAAATCCACATGAGAGAGAGAGAGAG 1304
Db      1291  TGGATCTTGTCTGACCTGAGCTGTCGAGCAAAATCCACATGAGAGAGAGAGAGAG 1350
QY      1305  TTGCTGCGACCTGATGTCAGCTCACTTCTCTGATGCTTCTGAGCCCTGGCTAC 1364
Db      1351  TTGCTGCGACCTGATGTCAGCTCACTTCTCTGATGCTTCTGAGCCCTGGCTAC 1410
QY      1365  ATCTCTGAGCCCTTAATACATGTCGTGGGTATCAACATGAGTGTGATGTCACAC 1424
Db      1411  ATCTCTGAGCCCTTAATACATGTCGTGGGTATCAACATGAGTGTGATGTCACAC 1470
QY      1425  GTGCTCAAGCAGAGAGACTTTGTGTCCATGCTGTGTAGAAAACAGACTGGGAGACC 1484
Db      1471  GTGCTCAAGCAGAGAGACTTTGTGTCCATGCTGTGTAGAAAACAGACTGGGAGACC 1530
QY      1485  TTATGTGAGCAGACATCCCAAGAGAGAGAGATGCTTCTTCTTCTTCTGATG 1544
Db      1531  TTATGTGAGCAGACATCCCAAGAGAGAGAGATGCTTCTTCTTCTTCTGATG 1590
QY      1545  TTCTGTCTGGGAGAGCTTCAAGAGATTTGTGGCTGAGAGGCTATTAAAGCAGACAG 1604
Db      1591  TTCTGTCTGGGAGAGCTTCAAGAGATTTGTGGCTGAGAGGCTATTAAAGCAGACAG 1650
QY      1605  TATCAGTGAATTTGATCCATAAAGCTCCCTGTCCACATTTGCCCAATGGGAGATGATC 1664
Db      1651  TATCAGTGAATTTGATCCATAAAGCTCCCTGTCCACATTTGCCCAATGGGAGATGATC 1710
QY      1665  TTTACAGAAAGAGCTACACAGATTTTCCACAGAGATGC 1703
Db      1711  TTTACAGAAAGAGCTACACAGATTTTCCACAGAGATGC 1749

```

RESULT 4
US-08-992-334-1

```

; Sequence 1, Application US/08992334
; Patent No. 5919678
; GENERAL INFORMATION:
; APPLICANT: Gruss, Alexandra
; TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
; TITLE OF INVENTION: PLASMID
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christie Parker & Hale, LLP
; STREET: 350 West Colorado Boulevard, Suite 500
; CITY: Pasadena
; STATE: California
; COUNTRY: United States
; ZIP: 91105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/992,334
; FILING DATE: 17-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,752
; FILING DATE: 24-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00248
; FILING DATE: 12-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/03034
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Priou, D. Bruce
; REGISTRATION NUMBER: 20958
; REFERENCE/DOCKET NUMBER: C93:31779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 795-9900
; TELEFAX: (626) 577-8800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3792 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: YES
; IMMEDIATE SOURCE:
; CLONE: pg+host4
; US-08-992-334-1

```

```

Query Match      0.9%; Score 19; DB 2; Length 3792;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1821  TCTCTGACAAAGAGCAAG 1839
Db      1818  TCTCTGACAAAGAGCAAG 1836

```

RESULT 5
US-08-302-752-1
; Sequence 1, Application US/08302752
; Patent No. 6025190
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: THERMOSENSITIVE PLASMID
; NUMBER OF SEQUENCES: 3
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3792 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-1

Query Match 0.9%; Score 19; DB 3; Length 3792;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCCTGCAAAAGAGCAAG 1839
 ||||||||||||||||
DB 1818 TCCTGCAAAAGAGCAAG 1836

RESULT 6
US-08-992-334-2
Sequence 2, Application US/08992334
Patent No. 5919678
GENERAL INFORMATION:
APPLICANT: Gruss, Alexandra
APPLICANT: Maguin, Emmanuelle
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
TITLE OF INVENTION: PLASMID
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christie Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR FR92/03034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Proulx, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93.31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 5234 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-992-334-2

Query Match 0.9%; Score 19; DB 2; Length 5234;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCCTGCAAAAGAGCAAG 1839
 ||||||||||||||||
DB 3260 TCCTGCAAAAGAGCAAG 3278

RESULT 7
US-08-302-752-2
Sequence 2, Application US/08302752
Patent No. 6025190
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: THERMOSENSIBLE PLASMID
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5234 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-2

Query Match 0.9%; Score 19; DB 3; Length 5234;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCCTGCAAAAGAGCAAG 1839
 ||||||||||||||||
DB 3260 TCCTGCAAAAGAGCAAG 3278

RESULT 8
US-08-992-334-3
Sequence 3, Application US/08992334
Patent No. 5919678
GENERAL INFORMATION:
APPLICANT: Gruss, Alexandra
APPLICANT: Maguin, Emmanuelle
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
TITLE OF INVENTION: PLASMID
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christie Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California

```

; COUNTRY: United States
; ZIP: 91105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; .APPLICATION NUMBER: US/08/992,334
; FILING DATE: 17-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,752
; FILING DATE: 24-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00248
; FILING DATE: 12-MAR-1993
; APPLICATION DATA:
; APPLICATION NUMBER: FR 992/03034
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Prout, D. Bruce
; REGISTRATION NUMBER: 20958
; REFERENCE/DOCKET NUMBER: C93:31779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 795-9900
; TELEFAX: (626) 577-8800
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6722 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; US-08-992-334-3

Query Match          0.9%; Score 19; DB 2; Length 6722;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1821 TCTCTGCAAAAGAGCAAG 1839
      |||||||||||||||
DB      4748 TCTCTGCAAAAGAGCAAG 4766

RESULT 9
; Sequence 3, Application US/08302752
; Patent No. 6025190
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: THERMOSENSIBLE PLASMID
; NUMBER OF SEQUENCES: 3
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,752
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9203034
; FILING DATE: 13-MAR-1992
; APPLICATION DATA:
; APPLICATION NUMBER: WO FR/93/00248
; FILING DATE: 12-MAR-1993
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6722 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

```

; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-302-752-3

Query Match          0.9%; Score 19; DB 3; Length 6722;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1821 TCTCTGCAAAAGAGCAAG 1839
      |||||||||||||||
DB      4748 TCTCTGCAAAAGAGCAAG 4766

RESULT 10
; US-09-439-313-320
; Sequence 320, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqul
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Ranger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 320
; LENGTH: 150
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-439-313-320

Query Match          0.9%; Score 18; DB 4; Length 150;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GAATTCATTGTTGGG 18
      |||||||||||||||
DB      31 GAATTCATTGTTGGG 48

RESULT 11
; US-09-352-616A-320
; Sequence 320, Application US/09352616A
; Patent No. 6395278
; GENERAL INFORMATION:
; APPLICANT: Dillon, Davin C.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang, Yuqul
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.427C8
; CURRENT APPLICATION NUMBER: US/09/352,616A
; CURRENT FILING DATE: 1999-07-13
; NUMBER OF SEQ ID NOS: 472
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 320
; LENGTH: 150
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-352-616A-320
```

Query Match 0.9%; Score 18; DB 4; Length 150;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCATGTGTGGG 18
|||||
DB 31 GAATTCATGTGTGGG 48

RESULT 12

US-09-232-149A-320
; Sequence 320, Application US/09232149A
; Patent No. 6465611
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, David C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
; FILE REFERENCE: 210121.427C6
; CURRENT APPLICATION NUMBER: US/09/232,149A
; CURRENT FILING DATE: 1999-01-15
; NUMBER OF SEQ ID NOS: 338
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 320
; LENGTH: 150
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-232-149A-320

Query Match 0.9%; Score 18; DB 4; Length 150;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCATGTGTGGG 18
|||||
DB 31 GAATTCATGTGTGGG 48

RESULT 13

US-09-020-956-78
; Sequence 78, Application US/09020956
; Patent No. 6261562
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, David C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020,956
; FILING DATE: 09-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 78:

SEQUENCE CHARACTERISTICS:

LENGTH: 201 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-09-020-956-78

Query Match 0.9%; Score 18; DB 3; Length 201;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCATGTGTGGG 18
|||||
DB 18 GAATTCATGTGTGGG 35

RESULT 14

US-09-030-607-78
; Sequence 78, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, David C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-09-030-607-78

Query Match 0.9%; Score 18; DB 3; Length 201;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCATGTGTGGG 18
|||||
DB 18 GAATTCATGTGTGGG 35

RESULT 15

US-09-439-313-78

```

? Sequence 78, Application US/09439313
? Patent No. 6329505
?
? GENERAL INFORMATION:
?
? APPLICANT: Xu, Jiangchun
? APPLICANT: Dillon, Davin C.
? APPLICANT: Mitcham, Jennifer L.
? APPLICANT: Harlocker, Susan Louise
? APPLICANT: Jiang Yuqi
? APPLICANT: Reed, Steven G.
? APPLICANT: Kalos, Michael
? APPLICANT: Fanger, Gary
? APPLICANT: Retter, Mark
? APPLICANT: Solk, John
?
? APPLICANT: Day, Craig
?
? TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
? TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
?
? FILE REFERENCE: 210121.42709
?
? CURRENT APPLICATION NUMBER: US/09/439,313
?
? NUMBER OF SEQ ID NOS: 575
?
? SOFTWARE: FastSeq for Windows Version 3.0
?
? SEQ ID NO 78
?
? LENGTH: 201
?
? TYPE: DNA
?
? ORGANISM: Homo sapien
?
? US-09-439-313-78

```

	0.9%:	Score	18;	DB	4;	Length	201;
Query Match							
Best Local Similarity	100.0%:	Pred.	No.	41,			
Matches	18;	Conservative	0;	Mismatches	0;	Indels	Gaps
Qy	1 GAATTCATTGTGTTGGG	18
Dq							
Db	GAATTCATTGTGTTGGG	35

Search completed: August 15, 2003, 04:16:10
Job time : 143 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2003, 00:56:07 ; Search time 1505 Seconds
(without alignments)
3037.948 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043

Sequence: 1 gaattccattgtgttgta.....tgggacccaataaaaaa 2043

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1504479 seqs, 1118970152 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2057348

Minimum DB seq length: 50

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1547	75.7	2032	US-09-816-825-1
2	1547	75.7	2032	US-10-007-262-2
3	403	19.7	505	US-09-998-598-2595
4	403	19.7	517	US-09-919-580-194
5	114	5.6	389	US-09-919-580-264
6	59	2.9	2544	US-09-927-602-1
7	59	2.9	48436	US-09-927-602-38
8	22	1.1	167	US-09-867-701-9458
9	22	1.1	171	US-09-867-701-9067
10	22	1.1	187	US-09-867-701-9185
11	22	1.1	191	US-09-867-701-9284
12	20	1.0	398	US-09-960-352-11099
13	20	1.0	570	US-10-027-632-140660
14	20	1.0	11102	US-10-205-823-334
15	20	1.0	180557	US-10-003-806-6
16	20	1.0	180557	US-10-003-806-9

C 17	19	0.9	341	10	US-09-867-701-8841	Sequence 8841, Ap
C 18	19	0.9	341	10	US-09-867-701-8888	Sequence 8888, Ap
C 19	19	0.9	349	9	US-09-759-143-649	Sequence 649, Ap
C 20	19	0.9	349	9	US-09-780-669-649	Sequence 649, Ap
C 21	19	0.9	349	9	US-09-822-827-649	Sequence 649, Ap
C 22	19	0.9	349	9	US-09-895-793-649	Sequence 649, Ap
C 23	19	0.9	349	9	US-09-895-814-649	Sequence 649, Ap
C 24	19	0.9	349	9	US-10-012-896-649	Sequence 649, Ap
C 25	19	0.9	405	10	US-09-867-701-9634	Sequence 9634, Ap
C 26	19	0.9	544	13	US-10-027-632-87347	Sequence 87347, A
C 27	19	0.9	766	13	US-10-027-632-118925	Sequence 118925,
C 28	19	0.9	2874	13	US-10-027-632-111785	Sequence 111785,
C 29	19	0.9	3273	12	US-09-971-392-26	Sequence 26, Ap
C 30	19	0.9	3274	13	US-10-002-600-552	Sequence 52, Ap
C 31	19	0.9	4234	19	US-09-815-242-9944	Sequence 9944, Ap
C 32	19	0.9	4537	13	US-10-108-605-322	Sequence 322, Ap
C 33	19	0.9	6346	12	US-10-017-161-2201	Sequence 2201, Ap
C 34	19	0.9	8742	9	US-09-908-711-147	Sequence 147, Ap
C 35	19	0.9	8742	11	US-09-764-891-6217	Sequence 6217, Ap
C 36	19	0.9	10929	14	US-10-032-393-7	Sequence 7, Ap
C 37	19	0.9	62944	10	US-09-954-456-2257	Sequence 2257, Ap
C 38	19	0.9	75899	9	US-09-854-883-243	Sequence 243, Ap
C 39	19	0.9	684973	10	US-09-263-959-1	Sequence 1, Ap
C 40	18	0.9	129	9	US-09-759-143-765	Sequence 765, App
C 41	18	0.9	129	9	US-09-780-669-765	Sequence 765, App
C 42	18	0.9	129	9	US-09-822-827-765	Sequence 765, App
C 43	18	0.9	129	10	US-09-895-793-765	Sequence 765, App
C 44	18	0.9	129	10	US-09-895-814-765	Sequence 765, App
C 45	18	0.9	129	13	US-10-012-896-765	Sequence 765, App

ALIGNMENTS

RESULT 1
US-09-816-825-1
Sequence 1, Application US/09816825
Patent No. US20010051370A1
GENERAL INFORMATION:
APPLICANT: Bistrup, Annette
APPLICANT: Rosen, Steven D.
APPLICANT: Hemmerich, Stefan
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
FILE REFERENCE: 6510-107CON
CURRENT APPLICATION NUMBER: US/09/816, 825
CURRENT FILING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: 09/045, 284
PRIOR FILING DATE: 1998-03-20
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2032
TYPE: DNA
ORGANISM: Homo sapiens
US-09-816-825-1

Query Match	75.7%	Score 1547	DB 9	Length 2032
Best Local Similarity	99.9%	Pred. No. 0		
Matches 1597	Conservative	0	Mismatches	1
			Indels	0
			Gaps	0
QY 106	CAAGGCTTCACCTTCAGCACAATGCTACGCTAAATAAATGAGTCTGCTTTTC	165		
DB 160	CAAGGCTTCACCTTCAGCACAATGCTACGCTAAATAAATGAGTCTGCTTTTC	219		
QY 166	GGTTTCACAGATGGCATCTTGGCTATCTTCACATGTACAGCCACAAATCAGCTC	225		
DB 220	GGTTTCACAGATGGCATCTTGGCTATCTTCACATGTACAGCCACAAATCAGCTC	279		
QY 226	CCGTCTATGAGGACGACGCGGATGACGCGTCTGCTCTTCTCGGCGCTC	285		
DB 280	CCGTCTATGAGGACGACGCGGATGACGCGTCTGCTCTTCTCGGCGCTC	339		
QY 286	TGGCTCTCTTTTGTGGGCGAGCTTTTGGGCGACGACCCAGATGTTTCTACTGATNGA	345		

Db 340 TGGCTCTTTTGGGGGAGCTTTTGGGAGCAGCCAGATGTTTCTACCTGATGGA 399
 QY 346 GCCCGCTGGCAGAGTGGATGACCTTCAAGCAGACAGCCGGCTGGATGCTGACATGGC 405
 Db 400 GCCCGCTGGCAGAGTGGATGACCTTCAAGCAGACAGCCGGCTGGATGCTGACATGGC 459
 QY 406 TGTGGGAGTCTGATACGGGCGCTTCTGTGGCAGATGAGCGTCTTGTATGCCATACAT 465
 Db 460 TGTGGGAGTCTGATACGGGCGCTTCTGTGGCAGATGAGCGTCTTGTATGCCATACAT 519
 QY 466 GGAACCTGGTCCCGGAGACAGTCCAGCTTTTCAAGTGGGAGAACAGCCGGGCGCTGTG 525
 Db 520 GGAACCTGGTCCCGGAGACAGTCCAGCTTTTCAAGTGGGAGAACAGCCGGGCGCTGTG 579
 QY 526 TTCTGACCTGGCTGTGATCATCCACAGATGAAATCATCCCGGGCTCACTGACG 585
 Db 580 TTCTGACCTGGCTGTGATCATCCACAGATGAAATCATCCCGGGCTCACTGACG 639
 QY 586 GCTCCTGTGACGACAGACCCCTTGTGAGGTGGTGAAGAGCCCTGCCGCTCTACAGCA 645
 Db 640 GCTCCTGTGACGACAGACCCCTTGTGAGGTGGTGAAGAGCCCTGCCGCTCTACAGCA 699
 QY 646 CGTGGGCTCAAGAGGAGTGGCTTCTTCAACCTGGACGCTTACCGGCTGCTGAAGA 705
 Db 700 CGTGGGCTCAAGAGGAGTGGCTTCTTCAACCTGGACGCTTACCGGCTGCTGAAGA 759
 QY 706 CCCCTCCCTCAACCTGGATATGCTGACACCTGGTCCGGAGCCCGGGCGCTGTCCGTTTC 765
 Db 760 CCCCTCCCTCAACCTGGATATGCTGACACCTGGTCCGGAGCCCGGGCGCTGTCCGTTTC 819
 QY 766 CCGAGAGCCGACAAAGGAGATCTCATGTGACATGCGATTTGTATGGGCGAGATGA 825
 Db 820 CCGAGAGCCGACAAAGGAGATCTCATGTGACATGCGATTTGTATGGGCGAGATGA 879
 QY 826 GCAAAAACCTCAAGAGGAGACCAACCTACTATGTGATGAGAGTCACTGGCCAAAGCCA 885
 Db 880 GCAAAAACCTCAAGAGGAGACCAACCTACTATGTGATGAGAGTCACTGGCCAAAGCCA 939
 QY 886 GCTGAGATCTACAGACATCCAGTCTGTGCCAAGGCCCTGACGAAAGCTTACCTGCT 945
 Db 940 GCTGAGATCTACAGACATCCAGTCTGTGCCAAGGCCCTGACGAAAGCTTACCTGCT 999
 QY 946 TGTGGGCTATGAGGAGCTGGCTGAGCCCTGTGGGCGCCAGACTTCCCGAATGATGAAT 1005
 Db 1000 TGTGGGCTATGAGGAGCTGGCTGAGCCCTGTGGGCGCCAGACTTCCCGAATGATGAAT 1059
 QY 1006 CGTGGGATGGAATTTCTGCCCATCTTCAAGCTGGTGCATTAACATACCCGAGGCAA 1065
 Db 1060 CGTGGGATGGAATTTCTGCCCATCTTCAAGCTGGTGCATTAACATACCCGAGGCAA 1119
 QY 1066 GGGCATGGTGAACACGCTTTCCACAAATGCCAGGATGCCCTTAATGTCTCCAGGC 1125
 Db 1120 GGGCATGGTGAACACGCTTTCCACAAATGCCAGGATGCCCTTAATGTCTCCAGGC 1179
 QY 1126 TTTGGGCTGGTCTTGGCCCTATGAAAAGTTTCTGACCTTCAAGAAAGCCCTGGGCGATGC 1185
 Db 1180 TTTGGGCTGGTCTTGGCCCTATGAAAAGTTTCTGACCTTCAAGAAAGCCCTGGGCGATGC 1239
 QY 1186 CATGAATTTGCTGGGCTACCGCAGCTCAGATCTGAACAAGAAAGAGAACTGTGTGCT 1245
 Db 1240 CATGAATTTGCTGGGCTACCGCAGCTCAGATCTGAACAAGAAAGAGAACTGTGTGCT 1299
 QY 1246 GGAATCTTCTGTCTACCTGAGCTGCTCCCTGAGCAAAATCACTAAGAGGTTGAGAAAGCTT 1305
 Db 1300 GGAATCTTCTGTCTACCTGAGCTGCTCCCTGAGCAAAATCACTAAGAGGTTGAGAAAGCTT 1359
 QY 1306 TGTGCGACCTGCTGACGCTCAGTCACTTCTCTGATGCTCTGACGCTTGGCTTACA 1365
 Db 1360 TGTGCGACCTGCTGACGCTCAGTCACTTCTCTGATGCTCTGACGCTTGGCTTACA 1419
 QY 1366 TGTGCGACCTTAACTACATGCTGTGGGTATCACACTGAGTGTGATGCTGTCCACAG 1425

Db 1420 TCTGTGAGCCTTAACATACATGCTGTGGGTATCACAGTGAAGTGTGTGCCACAG 1479
 QY 1426 TGTGTACAGAGAGAGACTTTGTGTGTCATGCTGTGTAGAAAACAGACTGGGAACCT 1485
 Db 1480 TGTGTACAGAGAGAGACTTTGTGTGTCATGCTGTGTGTGTAGAAAACAGACTGGGAACCT 1539
 QY 1486 TATGTGAGCAGACATCCACAGAGGAAACAGGGATATGCTCTTCTTTCTGTGATCT 1545
 Db 1540 TATGTGAGCAGACATCCACAGAGGAAACAGGGATATGCTCTTCTTTCTGTGATCT 1599
 QY 1546 TCTGTCTGTGGGAGACTTATGAGACTTGTGGCCCTGGAGAGCCCTATTAAGCAGACAGAT 1605
 Db 1600 TCTGTCTGTGGGAGACTTATGAGACTTGTGGCCCTGGAGAGCCCTATTAAGCAGACAGAT 1659
 QY 1606 ATCAGTGAATTAATCATTAACCTCCCTGTCCACATCTTGGCCAAATGGGAATGATCT 1665
 Db 1660 ATCAGTGAATTAATCATTAACCTCCCTGTCCACATCTTGGCCAAATGGGAATGATCT 1719
 QY 1666 TTTCACCAAGAGCTCACAGCAATTTTCCACAGAGATGC 1703
 Db 1720 TTTCACCAAGAGCTCACAGCAATTTTCCACAGAGATGC 1757

RESULT 2
 US-10-007-262-2
 ; Sequence 2, Application US/10007262
 ; Publication No. US20020164748A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bistrup, Annette
 ; APPLICANT: Rosen, Steven D.
 ; APPLICANT: Tangemann, Kirsten
 ; APPLICANT: Hemmerich, Stefan
 ; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
 ; FILE REFERENCE: 6510-107CIP
 ; CURRENT APPLICATION NUMBER: US/10/007,262
 ; PRIOR FILING DATE: 2001-11-08
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 2032
 ; TYPE: DNA
 ; ORGANISM: H. sapiens
 ; US-10-007-262-2

Query Match 75.7%; Score 1547; DB 13; Length 2032;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 106 CAAGGCTCTTCACCTTCAGCACAATGCTACTGCTTAAAAAATGAAGCTCTGCTGTCTT 165
 Db 160 CAAGGCTCTTCACCTTCAGCACAATGCTACTGCTTAAAAAATGAAGCTCTGCTGTCTT 219
 QY 166 GGTTCCTCCAGATGGGCATTTGGCTATCTTCCACATGTACAGCCAAATCAGGCTC 225
 Db 220 GGTTCCTCCAGATGGGCATTTGGCTATCTTCCACATGTACAGCCAAATCAGGCTC 279
 QY 226 CCTGTCTATGAAGGACACAGCCCGAGGCGATGACAGTGTGTTCTGTCTTCCCTGGCGCTC 285
 Db 280 CCTGTCTATGAAGGACACAGCCCGAGGCGATGACAGTGTGTTCTGTCTTCCCTGGCGCTC 339
 QY 286 TGGCTCTCTTTTGTGGGCGACCTTTTGGGAGACCCAGATGTTTCTTACCTGATGGA 345
 Db 340 TGGCTCTCTTTTGTGGGCGACCTTTTGGGAGACCCAGATGTTTCTTACCTGATGGA 399
 QY 346 GCCCGCTGGCAGAGTGGATGACCTTCAAGAGAGACCCGCTGGATGCTGACATGGC 405
 Db 400 GCCCGCTGGCAGAGTGGATGACCTTCAAGAGAGACCCGCTGGATGCTGACATGGC 459
 QY 406 TGTGGGAGTCTGATACGGGCGCTTCTTGTGGAGACATGAGCGCTTGTGATGCTTACAT 465
 Db 460 TGTGGGAGTCTGATACGGGCGCTTCTTGTGGAGACATGAGCGCTTGTGATGCTTACAT 519

APPLICANT: Pyle, Ruth
APPLICANT: Xu, Jiangchun
TITLE OF INVENTION: SECRETIST, HEATHER
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.552
CURRENT APPLICATION NUMBER: US/09/919,580
CURRENT FILING DATE: 2001-07-30
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 194
LENGTH: 517
TYPE: DNA
ORGANISM: Homo sapiens
US-09-919-580-194

Query Match 19.7%; Score 403; DB 10; Length 517;
Best Local Similarity 100.0%; Pred. No. 8.5e-203; Indels 0; Gaps 0;
Matches 403; Conservative 0; Mismatches 0;

QY 1301 GGGTTGGCCACCTGGTGTGAGCTCAGTCACTTTCTGTGAATGCTTGTGAGCCTTGC 1360
DB 13 GGGTTGGCCACCTGGTGTGAGCTCAGTCACTTTCTGTGAATGCTTGTGAGCCTTGC 72
QY 1361 CTACATCTGAGCCTTAATACTAGATGCTGTGGGTATACACAGAGTGTGTCTC 1420
DB 73 CTACATCTGAGCCTTAATACTAGATGCTGTGGGTATACACAGAGTGTGTCTC 132
QY 1421 ACACGTGCTCAACAGAGAGCTTTGTGTGATGCTGTGTAGAAAACAGACGTGGG 1480
DB 133 ACACGTGCTCAACAGAGAGAGCTTTGTGTGATGCTGTGTAGAAAACAGACGTGGG 192
QY 1481 AACCTATGTGACAGACATCCACAGTGAAGAGGATTTGCTCTTCTTCTTCTT 1540
DB 193 AACCTATGTGACAGACATCCACAGTGAAGAGGATTTGCTCTTCTTCTTCTT 252
QY 1541 GATCTCCGTGCTGGGAGAGCTCAGAGACTTGTGGCCGGAGGCTATTAGACAGAC 1600
DB 253 GATCTCCGTGCTGGGAGAGCTCAGAGACTTGTGGCCGGAGGCTATTAGACAGAC 312
QY 1601 ACAGATACAGTGAATGTATCAATAACCTCCCTGTCCACATTTGCCCATGGGAGATG 1660
DB 313 ACAGATACAGTGAATGTATCAATAACCTCCCTGTCCACATTTGCCCATGGGAGATG 372
QY 1661 GATCTTACCAAGAGCTCAGAGACTTTTCCACAGAGATGC 1703
DB 373 GATCTTACCAAGAGCTCAGAGACTTTTCCACAGAGATGC 415

RESULT 5
US-09-919-580-264
Sequence 264, Application US/09919580
Patent No. US20020110832A1
GENERAL INFORMATION:
APPLICANT: Pyle, Ruth
APPLICANT: Xu, Jiangchun
TITLE OF INVENTION: SECRETIST, HEATHER
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.552
CURRENT APPLICATION NUMBER: US/09/919,580
CURRENT FILING DATE: 2001-07-30
NUMBER OF SEQ ID NOS: 934
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 264
LENGTH: 389
TYPE: DNA
ORGANISM: Homo sapiens
US-09-919-580-264

Query Match 5.6%; Score 114; DB 10; Length 389;
Best Local Similarity 100.0%; Pred. No. 1.3e-49;
Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1930 TGCCATCTGTATACTATAAATCCCAATTAAGTGTCTGTAGAAATGTCCTTTTATG 1989
DB 266 TGCCATCTGTATACTATAAATCCCAATTAAGTGTCTGTAGAAATGTCCTTTTATG 325
QY 1990 CTCTTAATTAATTAAGAGTAATGTTCATTTTATGGATCTCTTAATTAATTAATTA 2043
DB 326 CTCTTAATTAATTAAGAGTAATGTTCATTTTATGGATCTCTTAATTAATTAATTA 379

RESULT 6
US-09-927-602-1
Sequence 1, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPLICANT: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2544
TYPE: DNA
ORGANISM: Homo Sapien
FEATURE:
NAME/KEY: CDS
LOCATION: (693)...(1877)
US-09-927-602-1

Query Match 2.9%; Score 59; DB 9; Length 2544;
Best Local Similarity 100.0%; Pred. No. 1.9e-20;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGCGCTCTCTACAGCAGCGTGTCTCAAGAGAGTGCCTTCTTCAACCTGCAG 682
DB 1180 AGGCTGCGCGCTCTCTACAGCAGCGTGTCTCAAGAGAGTGCCTTCTTCAACCTGCAG 1238

RESULT 7
US-09-927-602-38
Sequence 38, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPLICANT: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 38
LENGTH: 48436
TYPE: DNA
ORGANISM: Homo Sapien
US-09-927-602-38

Query Match 2.9%; Score 59; DB 9; Length 48436;
Best Local Similarity 100.0%; Pred. No. 1.9e-20;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGCGCTCTCTACAGCAGCGTGTCTCAAGAGAGTGCCTTCTTCAACCTGCAG 682
DB 48024 AGGCTGCGCGCTCTCTACAGCAGCGTGTCTCAAGAGAGTGCCTTCTTCAACCTGCAG 48082

RESULT 8
US-09-867-701-9458
; Sequence 9458, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9458
; LENGTH: 167
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9458

Query Match 1.1%; Score 22; DB 10; Length 167;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 GAAGACGACAGAGGCTAGAGG 44
Db 3 GAAGACGACAGAGGCTAGAGG 24

RESULT 9
US-09-867-701-9067
; Sequence 9067, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9067
; LENGTH: 171
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9067

Query Match 1.1%; Score 22; DB 10; Length 171;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 GAAGACGACAGAGGCTAGAGG 44
Db 9 GAAGACGACAGAGGCTAGAGG 30

RESULT 10
US-09-867-701-9185
; Sequence 9185, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497

; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9185
; LENGTH: 187
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9185

Query Match 1.1%; Score 22; DB 10; Length 187;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 GAAGACGACAGAGGCTAGAGG 44
Db 32 GAAGACGACAGAGGCTAGAGG 53

RESULT 11
US-09-867-701-9284
; Sequence 9284, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9284
; LENGTH: 191
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9284

Query Match 1.1%; Score 22; DB 10; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 GAAGACGACAGAGGCTAGAGG 44
Db 29 GAAGACGACAGAGGCTAGAGG 50

RESULT 12
US-09-960-352-11099/C
; Sequence 11099, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengping C.
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11099
; LENGTH: 398
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 48-BOVMS1-008-Q1-E1-DB
US-09-960-352-11099

Query Match 1.0%; Score 20; DB 10; Length 398;
Best Local Similarity 100.0%; Pred. No. 9;

Matches 20: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 GATGCCATCTGGCTCTAT 194
|||||
Db 240 GATGCCATCTGGCTCTAT 221

RESULT 13
US-10-027-632-140660
; Sequence 140660, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Mang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 140660
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-140660

Query Match
Best Local Similarity 100.0%; Score 20; DB 13; Length 570;
Pred. No. 9;
Matches 20: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 GAAAGACCCCTCCTCAACC 719
|||||
Db 273 GAAAGACCCCTCCTCAACC 292

RESULT 14
US-10-205-823-334/C
; Sequence 334, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Kanjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangl
; APPLICANT: Monsey, Angela M.
; APPLICANT: Glatte, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356

; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 334
; LENGTH: 11102
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-205-823-334

Query Match
Best Local Similarity 100.0%; Score 20; DB 14; Length 11102;
Pred. No. 8.9;
Matches 20: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1523 TGCTCTCTCTCTTTCTGA 1542
|||||
Db 722 TGCTCTCTCTCTTTCTGA 703

RESULT 15
US-10-003-806-6
; Sequence 6, Application US/10003806
; Publication No. US20020119929A1
; GENERAL INFORMATION:
; APPLICANT: Bishop, Collin E.
; APPLICANT: Agoulrik, Alexander I.
; APPLICANT: Zhu, Qichao
; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
; FILE REFERENCE: P02066051/10024824
; CURRENT APPLICATION NUMBER: US/10/003,806
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 60/245,872
; PRIOR FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 180557
; TYPE: DNA
; ORGANISM: Human
US-10-003-806-6

Query Match
Best Local Similarity 100.0%; Score 20; DB 13; Length 180557;
Pred. No. 8.7;
Matches 20: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1893 TTTTCCATTACATAGAAA 1912
|||||
Db 2567 TTTTCCATTACATAGAAA 2586

Search completed: August 15, 2003, 04:41:29
Job time : 1507 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2003, 03:08:38 ; Search time 7281 seconds
(without alignments)
11478.968 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043
Sequence: 1 gaattccatcgtgtggtga.....tggatcctaataaaaaa 2043

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 2045481386 residues

Word size: 0

Total number of hits satisfying chosen parameters: 3519286

Minimum DB seq length: 300

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

Genbank1:
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rnd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1722	84.3	1992	9	AF280088 Homo sapi
2	1653	80.9	2037	9	BC035282 Homo sapi
3	1598	78.2	2011	9	AK026635 Homo sapi
4	1597	78.2	183228	9	AC010547 Homo sapi
5	1597	78.2	206943	2	AC138848 Homo sapi
6	1547	75.7	1979	6	BD127258 Homo sapi
7	1547	75.7	1979	6	AK074746 Homo sapi
8	1547	75.7	2032	6	AK020335 Homo sapi
9	1547	75.7	2032	6	BD134772 Homo sapi
10	1547	75.7	2032	9	AF131235 Homo sapi
11	1262	61.8	1333	9	AF149783 Homo sapi
12	1153	56.4	2065	6	AR310485 Homo sapi
13	695	34.0	877	6	BD124832 Homo sapi
14	695	34.0	877	6	BD126950 Homo sapi
15	403	19.7	517	6	AX381256 Homo sapi
16	114	5.6	389	6	AX381326 Homo sapi
17	105	5.1	548	6	BD125617 Homo sapi
18	59	2.9	1462	9	AF176839 Homo sapi
19	59	2.9	1647	6	AX327230 Homo sapi
20	59	2.9	2170	9	AF176838 Homo sapi
21	59	2.9	2544	9	AF219990 Homo sapi
22	59	2.9	3278	9	AF246718 Homo sapi
23	59	2.9	3786	9	AF280086 Homo sapi
24	59	2.9	71503	9	AF219991 Homo sapi
25	59	2.9	157358	9	AC009163 Homo sapi
26	59	2.9	194832	9	AC025287 Homo sapi
27	59	2.9	208185	2	AC009105 Homo sapi
28	27	1.3	1926	6	BD134773 Homo sapi
29	27	1.3	1926	10	AF131236 Homo sapi
30	30	1.3	2201	10	AF109155 Homo sapi
31	27	1.3	232745	2	AC130174 Homo sapi
32	27	1.3	262679	2	AC110314 Homo sapi
33	24	1.2	228569	2	AC121740 Homo sapi
34	24	1.2	235173	2	AC133449 Homo sapi
35	24	1.2	240901	2	AC096328 Homo sapi
36	24	1.2	242253	2	AC095651 Homo sapi
37	37	1.2	246877	2	AC137365 Homo sapi
38	24	1.2	258097	2	AC129764 Homo sapi
39	24	1.2	271230	2	AC094919 Homo sapi
40	23	1.1	1740	10	AF176841 Homo sapi
41	23	1.1	1989	41	AF176840 Homo sapi
42	23	1.1	110000	2	AC099262 Homo sapi
43	23	1.1	159170	2	AC115914 Homo sapi
44	23	1.1	192365	2	AC139891 Homo sapi
45	23	1.1	194439	2	AC109022 Homo sapi

ALIGNMENTS

RESULT 1
AF280088
LOCUS Homo sapiens l-selectin ligand sulfotransferase GST-3 mRNA, complete cds.
ACCESSION AF280088
VERSION AF280088.1 GI:12060807
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1992)
Hemmerich, S., Lee, J.K., Bhakta, S., Bistrup, A., Ruddle, N.R. and Rosen, S.D.

FEATURES	source
TITLE	Chromosomal localization and genomic organization for the galactose-6-O-sulfotransferase gene family
JOURNAL	Glycobiology 11 (1), 75-87 (2001)
MEDLINE	2196027
PubMed	1181564
REFERENCE	2 (bases 1 to 1992)
AUTHORS	Hemmerich, S., Bhakta, S., Lee, J.-K., Bistrup, A., Ruddle, N.R. and Rosen, S.D.
TITLE	Direct Submission
JOURNAL	Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience, 3401 Hillview Ave., Palo Alto, CA 94304, USA
FEATURES	Location/Qualifiers
source	1..1992
gene	/organism="Homo sapiens"
CDS	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/chromosome="16"
	/map="16q23.1-q23.2"
	/cell_type="high endothelial"
	/class_type="lonsil"
	1..1992
	/gene="GST3"
	122..1282
	/gene="GST3"
	/note="similar to Homo sapiens sequence encoded by GenBank Accession Number AF131235"
	/codon_start=1
	/product="L-selectin ligand sulfotransferase GST-3"
	/protein_id="AAC84246.1"
	/db_xref="GI:1206808"
	/translation="MLPKRKILFLFVSOMATLALFPHYSHNISLSKAQPERHHVLISSRRSSSEYVQGLFGCHPDVFLMEPRMHWMTFKOSTAMLMHAYRLIRAVELCIDMSVDATMEPGPRROSSLPQWENSRLACSAPACDIIPDEILIPRAHRLCSQDPFEVEKACRSYSHVLEKEVEFNLOSYYPLIKPDSLMHLYHVRDPAVRSERERKGDLMIDRIWAGCHEOKLKKEDQPYVYMOYICOSLEIYTIOSLPALDERLYLVEYEDLARAVQATSRMYEVEGLFELPHLOIWHNTIRGKGDHAFHTNARALNYSQAMRMSLPYKVERLQACGDANMLLGRHVRBSQEQRNLLDLISLWYPEQIH"
BASE COUNT	482 a 540 c 489 g 481 t
ORIGIN	
Query Match	84.3%; Score 1722; DB 9; Length 1992;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1722; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	7 CATTGTGTGGGTACGGAAGACGACAGAAAGGTAGAGAGAGAAAAGCCATGGCCGGCTA 66
Db	1 CATTGTGTGGGTACGGAAGACGACAGAAAGGTAGAGAGAGAAAAGCCATGGCCGGCTA 60
QY	67 GCATGAGCCCTCTAAAAAGCAGCAGGGAAGGCCCAAGCCACAAGTCTTCCACTTCAGCAC 126
Db	61 GCATGAGCCCTCTAAAAAGCAGCAGGGAAGGCCCAAGCCACAAGTCTTCCACTTCAGCAC 120
QY	127 AATCTACTGCGCTAAAAAATGAAGTCACCTGTTTCTGGTTTCCAGATGGCCATCTT 186
Db	121 AATCTACTGCGCTAAAAAATGAAGTCACCTGTTTCTGGTTTCCAGATGGCCATCTT 180
QY	187 GCGCTATTCTTCCACATGTACACGCCAACATCAGCTCCCTGTCTATTGAAGCACAAGCC 246
Db	181 GCGCTATTCTTCCACATGTACACGCCAACATCAGCTCCCTGTCTATTGAAGCACAAGCC 240
QY	247 CGAGCGCATGACAGCTGCTGGTTCGTCTTCCCTGCGCCTGTGCGCTCTTCTTTTGGGGCA 306
Db	241 CGAGCGCATGACAGCTGCTGGTTCGTCTTCCCTGCGCCTGTGCGCTCTTCTTTTGGGGCA 300
QY	307 GCTTTTGGGAGAGACCCAGATGTTTCTACACGTATGAGACCCGCCCTGGGACAGCTGGAT 366
Db	301 GCTTTTGGGAGAGACCCAGATGTTTCTACACGTATGAGACCCGCCCTGGGACAGCTGGAT 360
QY	367 GACCTTCAGCAGACAGCCCGCTGGATGTGACATGGCTGTGGGGATCTGATACGGGC 426
Db	361 GACCTTCAGCAGACAGCCCGCTGGATGTGACATGGCTGTGGGGATCTGATACGGGC 420

OY	427	GGTCCTTCCTTGGCCGACATGAGCGTCCTTGTGATGCGCTACATATGGAAACCTGGTCCCGGGAGACA	486
Db	421	CGTCCTTCCTTGGCCGACATGAGCGTCCTTGTGATGCGCTACATATGGAAACCTGGTCCCGGGAGACA	480
OY	487	GTCGAGCCTCTTTCAGTGGGAGAACAGCCGGGCCCTGTGTTCTTCGACGCTGCCTGTGACAT	546
Db	481	GTCGAGCCTCTTTCAGTGGGAGAACAGCCGGGCCCTGTGTTCTTCGACGCTGCCTGTGACAT	540
OY	547	CATCCACAGATGAAATCATCCCCGGGGCTTCATGCAAGGCTCTGTGCATGCAACAGCC	606
Db	541	CATCCACAGATGAAATCATCCCCGGGGCTTCATGCAAGGCTCTGTGCATGCAACAGCC	600
OY	607	CTTTGAGGTGGTGGAGAGGCGCTGCCGCTCCTACAGCCACGCGTGTCTCAAGAGAGTGGC	666
Db	601	CTTTGAGGTGGTGGAGAGGCGCTGCCGCTCCTACAGCCACGCGTGTCTCAAGAGAGTGGC	660
OY	667	CTTCTTCAACCTGCAGTCCCTCTACACCGCTGCTGTAAGAACCCCTCCCTCAACCTGCATAT	726
Db	661	CTTCTTCAACCTGCAGTCCCTCTACACCGCTGCTGTAAGAACCCCTCCCTCAACCTGCATAT	720
OY	727	CGTGCACCTGGTCCGGGAGACCCCGGGCCGTGTTCGGTTCCGGAGAGCCACAAAGGAGAGA	786
Db	721	CGTGCACCTGGTCCGGGAGACCCCGGGCCGTGTTCGGTTCCGGAGAGCCACAAAGGAGAGA	780
OY	787	TCTCATGATTTACAGTGGCATTTGTGATGGGGAGCATATGAGCAAAACCTCAAGAAAGAGAGA	846
Db	781	TCTCATGATTTACAGTGGCATTTGTGATGGGGAGCATATGAGCAAAACCTCAAGAAAGAGAGA	840
OY	847	CCAAACCCCTACTATGTGATGACAGGTCATGTGCCAAAGCCAGCGTGAGATCTACAAACCAT	906
Db	841	CCAAACCCCTACTATGTGATGACAGGTCATGTGCCAAAGCCAGCGTGAGATCTACAAACCAT	900
OY	907	CCAGTCTTGGCCCAAGGCGCTGACAGAAACGCTACCTGTGTGCGCTATGAGAGCCTGGC	966
Db	901	CCAGTCTTGGCCCAAGGCGCTGACAGAAACGCTACCTGTGTGCGCTATGAGAGCCTGGC	960
OY	967	TCGAGCCCCCTGGGCCAGACACTTCCCGAATGTATTAATTCGtGGGATTTGGAAATTTTGTGCC	1026
Db	961	TCGAGCCCCCTGGGCCAGACACTTCCCGAATGTATTAATTCGtGGGATTTGGAAATTTTGTGCC	1020
OY	1027	CCATCTTCAGACCTGGGTGCTAAATCATACCCGAGGCAAGGGCATGGGTGACACAGCTT	1088
Db	1021	CCATCTTCAGACCTGGGTGCTAAATCATACCCGAGGCAAGGGCATGGGTGACACAGCTT	1088
OY	1087	CCACACAATGCCAGAGGATGCCCTTAATGTCTCCAGGCGTGGCGCTGGTCTTGGCCCTA	1146
Db	1081	CCACACAATGCCAGAGGATGCCCTTAATGTCTCCAGGCGTGGCGCTGGTCTTGGCCCTA	1140
OY	1147	TGAAAAGGTTTCTGCATCTTCAGAAAAGCCTGTGGCGCATATCCATGAATTTGCTGGGCTACCG	1206
Db	1141	TGAAAAGGTTTCTGCATCTTCAGAAAAGCCTGTGGCGCATATCCATGAATTTGCTGGGCTACCG	1200
OY	1207	CCAGGTCAGATCTGAAACAAGAACAGAGAAACCTGTGTGTGATCTTCTGTCTACCTGGAC	1266
Db	1201	CCAGGTCAGATCTGAAACAAGAACAGAGAAACCTGTGTGTGATCTTCTGTCTACCTGGAC	1266
OY	1267	TGTCCCTGAGCAAAATCCACTAAGAGGGTTGAAAGGCTTTGCTGCGACACTGGTGTCAAGC	1326
Db	1261	TGTCCCTGAGCAAAATCCACTAAGAGGGTTGAAAGGCTTTGCTGCGACACTGGTGTCAAGC	1320
OY	1327	TCAGTCACTTCTCTGAAATGCTTTCAGAGCCTTGGCTACATCTCTGAGCCTTAACATCAATG	1386
Db	1321	TCAGTCACTTCTCTGAAATGCTTTCAGAGCCTTGGCTACATCTCTGAGCCTTAACATCAATG	1380
OY	1387	TCTGTGGGTATCACAGTGAAGTGTGATGTGTCTCACACGTGCTCAAGCAGAAAGGACTTT	1446
Db	1381	TCTGTGGGTATCACAGTGAAGTGTGATGTGTCTCACACGTGCTCAAGCAGAAAGGACTTT	1440
OY	1447	GTCGTCATGCTTGTGTGTAGAAAACAGACTGGGAAACCTTATGTGAGCAGACATCCAC	1508
Db	1441	GTCGTCATGCTTGTGTGTGTAGAAAACAGACTGGGAAACCTTATGTGAGCAGACATCCAC	1500
OY	1507	CAGTGAACAGAGGATATGTCTCTTCTTCTTGTGATCTTCCTGTGTGGGAGACTTCAG	1566

QY 951 GCTATAGAGACCTGGCTCGAGCCCTGTGGCCAGACTTCCGAATGATGAATTCGTGG 1010
 DB 901 GCTATAGAGACCTGGCTCGAGCCCTGTGGCCAGACTTCCGAATGATGAATTCGTGG 960
 QY 1011 GATTGGAATTTCTGGCCCATCTTCAGACCTGGGTGCATATACACCCAGGGAAGGCA 1070
 DB 961 GATTGGAATTTCTGGCCCATCTTCAGACCTGGGTGCATATACACCCAGGGAAGGCA 1020
 QY 1071 TGGGTGACCAAGCTTTCCACACAAATGCCAGGATGCCCTTATATGCTCCAGGCTTGGC 1130
 DB 1021 TGGGTGACCAAGCTTTCCACACAAATGCCAGGATGCCCTTATATGCTCCAGGCTTGGC 1080
 QY 1131 GCTGCTCTTGGCCCTTGAAGAGGCTTCTGCACTTGAAGAGGCTTGGGATGCAATGA 1190
 DB 1081 GCTGCTCTTGGCCCTTGAAGAGGCTTCTGCACTTGAAGAGGCTTGGGATGCAATGA 1140
 QY 1191 ATTGCTGGGCTTCGGCCAGCTGATGATGAAACAAACAGAGAAACCTTCTCTGATC 1250
 DB 1141 ATTGCTGGGCTTCGGCCAGCTGATGATGAAACAAACAGAGAAACCTTCTCTGATC 1200
 QY 1251 TTCTGTACTGAGTGTGCTGCTGAGCAATTCACATAGAGGCTTGAAGAGCTTGGCTG 1310
 DB 1201 TTCTGTACTGAGTGTGCTGCTGAGCAATTCACATAGAGGCTTGAAGAGCTTGGCTG 1260
 QY 1311 CCACCTGCTGTACAGCTCAGTCACTTCTCTGATGAGCTTGGCTGATCTCT 1370
 DB 1261 CCACCTGCTGTACAGCTCAGTCACTTCTCTGATGAGCTTGGCTGATCTCT 1320
 QY 1371 GAGCCTTAACATGCTGTGGGTATGACACATGATGATGATGATGATGATGATGATGATG 1430
 DB 1321 GAGCCTTAACATGCTGTGGGTATGACACATGATGATGATGATGATGATGATGATGATG 1380
 QY 1431 AAGCAGAGAGACTTTTGTGTGTCATGCTTGTGTAGAAAACAGACAGTGGGAACTTATGT 1490
 DB 1381 AAGCAGAGAGACTTTTGTGTGTCATGCTTGTGTAGAAAACAGACAGTGGGAACTTATGT 1440
 QY 1491 GAGCAGACATCCACACATGAGAAAGGATGATGCTTCTTCTTCTTCTTCTTCTTCTTCT 1550
 DB 1441 GAGCAGACATCCACACATGAGAAAGGATGATGCTTCTTCTTCTTCTTCTTCTTCTTCT 1500
 QY 1551 TCTGGCAGACTTCAGAGACTTGTGGCTGAGAGGCTTATTAAGCAGACAGATGATGATG 1610
 DB 1501 TCTGGCAGACTTCAGAGACTTGTGGCTGAGAGGCTTATTAAGCAGACAGATGATGATG 1560
 QY 1611 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1670
 DB 1561 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1620
 QY 1671 CAAGAGCTCAGCAGACTTTCACAGAGATGC 1703
 DB 1621 CAAGAGCTCAGCAGACTTTCACAGAGATGC 1653

RESULT 3
 AK026635
 LOCUS Homo sapiens cdna: FLJ22982 fls, clone KAT11454, highly similar to AF131235 Homo sapiens N-acetylglyucosamine 6-O-sulfotransferase mRNA.
 DEFINITION
 AK026635
 VERSION AK026635.1 GI:10439531
 KEYWORDS oligo capping, fls (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (sites)
 Matanabe, K., Kumagai, A., Itakura, S., Yamazaki, M., Tashiro, H., Ota, T., Suzuki, Y., Ohashi, M., Nishi, T., Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.
 NEDO human cdna sequencing project
 Unpublished
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 2 (bases 1 to 2011)

AUTHORS Sugano, S., Suzuki, Y., Ota, T., Ohashi, M., Nishi, T., Isogai, T., Shibahara, T., Tanaka, T. and Nakamura, Y.
 TITLE Direct Submission
 JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail: cdna@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286, Fax: 81-3-5449-5416)
 COMMENT NEDO human cdna sequencing project supported by Ministry of International Trade and Industry of Japan: cdna full insert sequencing: Research Association for Biotechnology; cdna library construction, 5' - 6' 3' end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
 FEATURES
 source location/Qualifiers
 1..2011
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="KAT11454"
 /cell_line="KATO III"
 /cell_type="signet-ring cell carcinoma"
 /note="cloning vector pME18SFL3"
 1..2011
 /note="highly similar to AF131235 Homo sapiens N-acetylglyucosamine 6-O-sulfotransferase mRNA"
 BASE COUNT 482 a 541 c 475 g 513 t
 ORIGIN
 Query Match 78.2%; Score 1598; DB 9; Length 2011;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1598; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 106 CAAGGCTTCCATTTAGCAGCAATGCTACGCTTAATAAATAGCTCTGCTTTCT 165
 DB 50 CAAGGCTTCCATTTAGCAGCAATGCTACGCTTAATAAATAGCTCTGCTTTCT 109
 QY 166 GATTTCAGAGATGGCATCTTGGCTATCTTCCACATGATGACGACACAACTAGCCT 225
 DB 110 GATTTCAGAGATGGCATCTTGGCTATCTTCCACATGATGACGACACAACTAGCCT 169
 QY 226 CCTGTATGAGGACACAGCCGAGCATGACGCTGTCTCTCTCTGCGCTC 285
 DB 170 CCTGTATGAGGACACAGCCGAGCATGACGCTGTCTCTCTCTGCGCTC 229
 QY 286 TGCTCTCTCTTGTGGGAGCTTTTGGGAGCAACCAAGTCTTCTACTGATGGA 345
 DB 230 TGCTCTCTCTTGTGGGAGCTTTTGGGAGCAACCAAGTCTTCTACTGATGGA 289
 QY 346 GCCCGCTGACAGTGTGATGACCTTCAAGCAGACAGCCGCTGATGCTGACATGCG 405
 DB 290 GCCCGCTGACAGTGTGATGACCTTCAAGCAGACAGCCGCTGATGCTGACATGCG 349
 QY 406 TGTGCGGAGATGATACGGGCGCTCTTCTTGTGCGACATGAGCGCTTTGATGCTTACAT 465
 DB 350 TGTGCGGAGATGATACGGGCGCTCTTCTTGTGCGACATGAGCGCTTTGATGCTTACAT 409
 QY 466 GGAACCTGGTCCCGGAGACAGTCCAGCCCTTCAAGTGGGAGAAAGCGCGGCGCTG 525
 DB 410 GGAACCTGGTCCCGGAGACAGTCCAGCCCTTCAAGTGGGAGAAAGCGCGGCGCTG 469
 QY 526 TTCTGCACTGCTGTGATCATCTCCACAAATGAAATCATCCCGGCGCTCACTGAC 585
 DB 470 TTCTGCACTGCTGTGATCATCTCCACAAATGAAATCATCCCGGCGCTCACTGAC 529
 QY 586 GCTCTGTGCAAGTCAACAGCCCTTGAAGTGTGGAAGAGCGCTGCGCTCTTACAGCA 645
 DB 530 GCTCTGTGCAAGTCAACAGCCCTTGAAGTGTGGAAGAGCGCTGCGCTCTTACAGCA 589
 QY 646 CGTGGCTCAAGAGAGTGGCTTTCACAGCTGATGCTTACCGCTGCTGAAGA 705


```

Db 590 CGTGTGCTCAAGAGAGTGCCTTCTTCAACCTGACAGTCCCTTACCCTGCTGAAGA 649
Qy 706 CCCCTCCCTCAACCTGCATATCGACACCTGTCGGGAGACCCCGGGCGCTGTCCGTC 765
Db 650 CCCCTCCCTCAACCTGCATATCGACACCTGTCGGGAGACCCCGGGCGCTGTCCGTC 709
Qy 766 CCGAAGACGCAAGAGAGATCTCATGATGACAGTCGACATTTGATGGGCGACATGA 825
Db 710 CCGAAGACGCAAGAGAGATCTCATGATGACAGTCGACATTTGATGGGCGACATGA 769
Qy 826 GCAAAAACCTCAAGAGAGAGACCAACCTTACTATGATGACAGTCATCTCCAAAGCA 885
Db 770 GCAAAAACCTCAAGAGAGAGACCAACCTTACTATGATGACAGTCATCTCCAAAGCA 829
Qy 886 GCTGAGATCTACAAAGACCATTCAGTCCCTGCCAAGGCGCTGAGAGAGAGCTACCTGCT 945
Db 830 GCTGAGATCTACAAAGACCATTCAGTCCCTGCCAAGGCGCTGAGAGAGAGCTACCTGCT 889
Qy 946 TGTGCGGTATGAGAGACCTGCTGAGCCCTGTGGCCAGACTTCCCGAATGTATGATTT 1005
Db 890 TGTGCGGTATGAGAGACCTGCTGAGCCCTGTGGCCAGACTTCCCGAATGTATGATTT 949
Qy 1006 CGTGGGATGGAATTTCTTGGCCCATCTTCAACACCTGGGTGATTAACATCACCCGAGCAA 1065
Db 950 CGTGGGATGGAATTTCTTGGCCCATCTTCAACACCTGGGTGATTAACATCACCCGAGCAA 1009
Qy 1066 GGGCATGGGTGACACGCTTTCACACAAATGCCAGGAGTCCCTTATGTCCTCCAGGC 1125
Db 1010 GGGCATGGGTGACACGCTTTCACACAAATGCCAGGAGTCCCTTATGTCCTCCAGGC 1069
Qy 1126 TTGGCGGTGCTTGTGGCCCATGAAAGGTTCTGACATTCAGAAAGCCTGTGGCGATGC 1185
Db 1070 TTGGCGGTGCTTGTGGCCCATGAAAGGTTCTGACATTCAGAAAGCCTGTGGCGATGC 1129
Qy 1186 CATGAATTTGCTGGGCTACACGCTTTCACACAAATGCCAGGAGTCCCTTATGTCCTCCAGGC 1245
Db 1130 CATGAATTTGCTGGGCTACACGCTTTCACACAAATGCCAGGAGTCCCTTATGTCCTCCAGGC 1189
Qy 1246 GGATCTTCTGCTACTGATGATGTCCTGAGCAATTCACCTAAGAGGGTTAGAAAGGCTT 1305
Db 1190 GGATCTTCTGCTACTGATGATGTCCTGAGCAATTCACCTAAGAGGGTTAGAAAGGCTT 1249
Qy 1306 TGTGCGACCTGCTGTGACGCTTCACTGATGATGCTTCTGAAAGCCTGTGGCGATGC 1365
Db 1250 TGTGCGACCTGCTGTGACGCTTCACTGATGATGCTTCTGAAAGCCTGTGGCGATGC 1309
Qy 1366 TCTGTGAGCCTTACTATGATGTCGTGGGTATCACACGATGATGTCCTCCAGGC 1425
Db 1310 TCTGTGAGCCTTACTATGATGTCGTGGGTATCACACGATGATGTCCTCCAGGC 1369
Qy 1426 TGTCAAGACAGAGAGACTTGTGTCATGCTGTGTCATGAAACAGAGCTGGGGAACCT 1485
Db 1370 TGTCAAGACAGAGAGACTTGTGTCATGCTGTGTCATGAAACAGAGCTGGGGAACCT 1429
Qy 1486 TATGTGAGCAGACATCCACAGTGAAGAGGGTATGTCCTTCTTCTTCTGATGCT 1545
Db 1430 TATGTGAGCAGACATCCACAGTGAAGAGGGTATGTCCTTCTTCTTCTGATGCT 1489
Qy 1546 TCTGTGCTGGGCACTTCAAGAGACTTGTGGCTGAGGCGCTATTAAGACGACAGAGCT 1605
Db 1490 TCTGTGCTGGGCACTTCAAGAGACTTGTGGCTGAGGCGCTATTAAGACGACAGAGCT 1549
Qy 1606 ATCAGTGAATTAAGATTAAGCTCCCTGTCACATGTCCTCCCAATGGGGAATGAGACT 1665
Db 1550 ATCAGTGAATTAAGATTAAGCTCCCTGTCACATGTCCTCCCAATGGGGAATGAGACT 1609
Qy 1666 TTCACAAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
Db 1610 TTCACAAAGAGCTCACAGCATTTTCCACAGAGATGC 1647

```

```

LOCUS AC010547 183228 bp DNA linear PRI 29-SEP-2001
DEFINITION Homo sapiens chromosome 16 clone RP11-510M2, complete sequence.
ACCESSION AC010547
VERSION AC010547.9 GI:15808510
KEYWORDS HTG.
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 183228)
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Human Chromosome 16
REFERENCE
AUTHORS 2 (bases 1 to 183228)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission.
REFERENCE
AUTHORS 3 (bases 1 to 183228)
TITLE DOE Joint Genome Institute.
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
This entry has been annotated with sequence quality
estimates computed by the phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
FEATURES
source
1..183228
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-510M2"
BASE COUNT 51901 a 42098 c 40026 g 49203 t
ORIGIN
Query Match 78.2%; Score 1597; DB 9; Length 183228;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 107 AAGCTCTTCACCTTACAGACAAATGCTACTGCTTAAAAAATGAAGCTCTCTGTTCTG 166
Db 142584 AAGGCTTTCACCTTACAGACAAATGCTACTGCTTAAAAAATGAAGCTCTCTGTTCTG 142643
Qy 167 GTTTCGCCAGATGGGCATTTGGCTCTATTTCCACATGTCACAGCCACAAATACAGCTCC 226
Db 142644 GTTTCGCCAGATGGGCATTTGGCTCTATTTCCACATGTCACAGCCACAAATACAGCTCC 142703
Qy 227 CTGTCTATGAAGGACACAGCCGAGGAGCATGACAGTGTGCTGTCTTCTGCGCGCTCT 286
Db 142704 CTGTCTATGAAGGACACAGCCGAGGAGCATGACAGTGTGCTGTCTTCTGCGCGCTCT 142763
Qy 287 GGCCTCTCTTTTGTGGGCGACCTTTTGGGCGACACCCAGATGTTTCTACCTGATGGAG 346
Db 142764 GGCCTCTCTTTTGTGGGCGACCTTTTGGGCGACACCCAGATGTTTCTACCTGATGGAG 142823

```



```

*      3767      5167: contig of 1401 bp in length
*      5168      5267: gap of unknown length
*      5268      12818: contig of 7551 bp in length
*      12819      36168: gap of unknown length
*      12919      36168: contig of 23251 bp in length
*      36170      36269: gap of unknown length
*      36270      206943: contig of 170674 bp in length.
      Location/Qualifiers
        source
          1. 206943
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone="RP11-1301B21"
            /clone_lib="RP11 human BAC library 11"
BASE COUNT   56312 a 44122 c 46163 g 59225 t 1121 others
ORIGIN
Query Match      78.2%; Score 1597; DB 2; Length 206943;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
      Oy      107  AAGCTTCCACCTTACAGACAAATGCTACCTGCTTAAAAAATGAAGCTCCTGCTTCTG 166
      Db      125368 AAGCTTCCACCTTACAGACAAATGCTACCTGCTTAAAAAATGAAGCTCCTGCTTCTG 125309
      Oy      167  GTTCCAGATGGCATCTGGCTCTATCTTCCACATGTACAGCCAAACATACGCTCC 226
      Db      125308 GTTCCAGATGGCATCTGGCTCTATCTTCCACATGTACAGCCAAACATACGCTCC 125249
      Oy      227  CTGTCTATGAAGACACAGCCGAGCATGCGATGCGTGTCTGTCTCTCTCTGCGCTCT 286
      Db      125248 CTGTCTATGAAGACACAGCCGAGCATGCGATGCGTGTCTGTCTCTCTGCGCTCT 125189
      Oy      287  GGCCTCTCTTTTGTGGGACAGCTTTTGGGACGACCCAGATGTTTCTACCTGATGAG 346
      Db      125188 GGCCTCTCTTTTGTGGGACAGCTTTTGGGACGACCCAGATGTTTCTACCTGATGAG 125129
      Oy      347  CCGGCTGGGACGCTGGATGACCTTCAAGACAGACCCGCTGGATGCTGACATGGCT 406
      Db      125128 CCGGCTGGGACGCTGGATGACCTTCAAGACAGACCCGCTGGATGCTGACATGGCT 125069
      Oy      407  GTGGGGATCTGATACGGGCGCTCTTGTGCGACATGAGACGCTTGTGATGCTACATG 466
      Db      125068 GTGGGGATCTGATACGGGCGCTCTTGTGCGACATGAGACGCTTGTGATGCTACATG 125009
      Oy      467  GAACCTGTCTCCCGGAGACAGTCCAGCCTCTTCAGTGGAGAAACAGCCGGCCCTGTGT 526
      Db      125008 GAACCTGTCTCCCGGAGACAGTCCAGCCTCTTCAGTGGAGAAACAGCCGGCCCTGTGT 124949
      Oy      527  TCTGACCTGCTGTGACATATCCCAAGATGAATCATCCCGGGGCTCACTGCAGG 586
      Db      124948 TCTGACCTGCTGTGACATATCCCAAGATGAATCATCCCGGGGCTCACTGCAGG 124889
      Oy      587  CTCCTGTGAGTCAACAGCCCTTGTGAGTGTGGAGAGGCTCCCTCTTACAGCCAC 646
      Db      124888 CTCCTGTGAGTCAACAGCCCTTGTGAGTGTGGAGAGGCTCCCTCTTACAGCCAC 124829
      Oy      647  GTGGTGTCTCAAGAGGTGGCTTCTTCAACCTGCAGTCCCTCAACCCGCTGTGAAGAC 706
      Db      124828 GTGGTGTCTCAAGAGGTGGCTTCTTCAACCTGCAGTCCCTCAACCCGCTGTGAAGAC 124769
      Oy      707  CCCGCCCTCAACGCGATATGTGACACCTGTGCGGGAGCCCGGGGCGCTTCCGCTTC 766
      Db      124768 CCCGCCCTCAACGCGATATGTGACACCTGTGCGGGAGCCCGGGGCGCTTCCGCTTC 124709
      Oy      767  CGAAGACGCAAAAGGAGATCTCATGTTGACAGTGCATGTTGATGGGACACATGAG 826
      Db      124708 CGAAGACGCAAAAGGAGATCTCATGTTGACAGTGCATGTTGATGGGACACATGAG 124649
      Oy      827  CAAAAAAGCAAGAGGAGGACCAACCTCTATATGTATGACAGGCTATCTGCCAAAGCCAG 886
      Db      124648 CAAAAAAGCAAGAGGAGGACCAACCTCTATATGTATGACAGGCTATCTGCCAAAGCCAG 124589

```

```

      Oy      887  CTGAGATCTTCAAGACATCCAGTCCCTGGCCCAAGGCCCTGCAGAACGCTTACCTGCTT 946
      Db      124588 CTGAGATCTTCAAGACATCCAGTCCCTGGCCCAAGGCCCTGCAGAACGCTTACCTGCTT 124529
      Oy      947  GTGGCTATGAGGACCTGGCTGAGCCCTGTGGCCAGACATTCGCAATGTATGAATTC 1006
      Db      124528 GTGGCTATGAGGACCTGGCTGAGCCCTGTGGCCAGACATTCGCAATGTATGAATTC 124469
      Oy      1007  GTGGATTTGGAATTTCTGGCCCATCTTCAAGACCTGGGTGCTATACATTCACCCGAGCAG 1066
      Db      124468 GTGGATTTGGAATTTCTGGCCCATCTTCAAGACCTGGGTGCTATACATTCACCCGAGCAG 124409
      Oy      1067  GGCATGGGTGACACAGGCTTTCACACAAATGCGAGGATGCGCTTAATGTCTCCAGGCT 1126
      Db      124408 GGCATGGGTGACACAGGCTTTCACACAAATGCGAGGATGCGCTTAATGTCTCCAGGCT 124349
      Oy      1127  TGGCGTGTCTTTCCTTATGAAAGGTTTCTGACTTCAGAAAGCCTGTGGCGATGCC 1186
      Db      124348 TGGCGTGTCTTTCCTTATGAAAGGTTTCTGACTTCAGAAAGCCTGTGGCGATGCC 124289
      Oy      1187  ATGAATTTGCTGGGCTACCGCCAGCTCAGATCTGAACAAAGACAGAAACCTTGTCTG 1246
      Db      124288 ATGAATTTGCTGGGCTACCGCCAGCTCAGATCTGAACAAAGACAGAAACCTTGTCTG 124229
      Oy      1247  GATCTGTCTACCTGGACTGTCCCTGAGCAATTCACCTAAGAGGGTTGAAGGCTTT 1306
      Db      124228 GATCTGTCTACCTGGACTGTCCCTGAGCAATTCACCTAAGAGGGTTGAAGGCTTT 124169
      Oy      1307  GCTGCCACCTGGTGTCAAGCCTCAGTCACTTCTCTGATGCTTGTGAGCCTTGCTACAT 1366
      Db      124168 GCTGCCACCTGGTGTCAAGCCTCAGTCACTTCTCTGATGCTTGTGAGCCTTGCTACAT 124109
      Oy      1367  CTCTGAGCCTTAATCAATGCTGTGGGTATCACAGTGAAGTGTGATGTTGTCTCCACAGT 1426
      Db      124108 CTCTGAGCCTTAATCAATGCTGTGGGTATCACAGTGAAGTGTGATGTTGTCTCCACAGT 124049
      Oy      1427  GCTCAGAGAGAGAGACTTTTGTGCATGCTTGTGTAGAAAACAGACTGGGAGACCTT 1486
      Db      124048 GCTCAGAGAGAGAGACTTTTGTGCATGCTTGTGTGTAGAAAACAGACTGGGAGACCTT 123989
      Oy      1487  ATGTGACAGACATCCACCAGATGAAACAGAGGATGCTTCTTCTTCTGTGATCTT 1546
      Db      123988 ATGTGACAGACATCCACCAGATGAAACAGAGGATGCTTCTTCTTCTGTGATCTT 123929
      Oy      1547  CCTGTCTGGGACACTTCAGAGACTTGTGGCTGTGGAGGCTATTAAAGACAGACAGTA 1606
      Db      123928 CCTGTCTGGGACACTTCAGAGACTTGTGGCTGTGGAGGCTATTAAAGACAGACAGTA 123869
      Oy      1607  TCAGTGGAAATGATCCATTAACCTCCCGTGCACATCTTGCCCAATGGGGAATGATCTT 1666
      Db      123868 TCAGTGGAAATGATCCATTAACCTCCCGTGCACATCTTGCCCAATGGGGAATGATCTT 123809
      Oy      1667  TCACCAAGAGCTCACCAGCATTTTTCACAGAGATGC 1703
      Db      123808 TCACCAAGAGCTCACCAGCATTTTTCACAGAGATGC 123772

RESULT 6
LOCUS      BD127258      1979 bp      DNA      linear      PAT 18-SEP-2002
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION  BD127258
VERSION    BD127258.1 GI:23222203
KEYWORDS   JP 2002017375-A/2689.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 1979)
            Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
            Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
            Koga,H.

```

TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002017375-A 2689 22-JAN-2002;
COMMENT HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002017375-A/2689
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253172
PI TOSHIO OTH, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
PI ISHII,
PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TERSUJI OTSUKI, HISASHI KOGA
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/PC
10, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
(69) . (1226).
FT CDS Location/Qualifiers
1. 1979
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 451 a 542 c 474 g 512 t
ORIGIN
Query Match 75.7%; Score 1547; DB 6; Length 1979;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; gaps 0;

647 CCCCTCCCTCAACCTCATATCGTCAGCCCTGTCGGGACCCCGGCGGTGTCGGTTC 706
QY
QY 766 CCGAGAACGCACAAAGAGATCTCATGATTCAGATCCCATGTCGATGGGAGCATGA 825
Db 707 CCGAGAACGCACAAAGAGATCTCATGATTCAGATCCCATGTCGATGGGAGCATGA 766
QY 826 GCAAAAACCTCAGAGAGAGACCAACCTACTATGATGATGACAGTCATCTGCCAAGCA 885
Db 767 GCAAAAACCTCAGAGAGAGACCAACCTACTATGATGATGACAGTCATCTGCCAAGCA 826
QY 886 GCTGGAGATCTACAACCATCCAGTCCTTGGCCCAAGGCGCTGACAGAGCCGACTGCT 945
Db 827 GCTGGAGATCTACAACCATCCAGTCCTTGGCCCAAGGCGCTGACAGAGCCGACTGCT 886
QY 946 TGTGCGCTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTTCCCGAATGATGAAT 1005
Db 887 TGTGCGCTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTTCCCGAATGATGAAT 946
QY 1006 CGTGGGATTTGGAATTTCTTGGCCCATCTTCAGACCTGGTGATATACATCACCGAGGCA 1065
Db 947 CGTGGGATTTGGAATTTCTTGGCCCATCTTCAGACCTGGTGATATACATCACCGAGGCA 1006
QY 1066 GGGCATGGGTGACACGCTTTCACACAAATGCGAGGATGACCTTAATGTCTCCAGGC 1125
Db 1007 GGGCATGGGTGACACGCTTTCACACAAATGCGAGGATGACCTTAATGTCTCCAGGC 1066
QY 1126 TTGGCGCTGCTTTTCCCTATGAAAAAGTTTCTGACTTCAGAAAGCCTGTGGCATGC 1185
Db 1067 TTGGCGCTGCTTTTCCCTATGAAAAAGTTTCTGACTTCAGAAAGCCTGTGGCATGC 1126
QY 1186 CATGAATTTGCTGGGTACCGCCAGCTGATGAAAGAAACAGAAAGAACTGTGCT 1245
Db 1127 CATGAATTTGCTGGGTACCGCCAGCTGATGAAAGAAACAGAAAGAACTGTGCT 1186
QY 1246 GGATCTTCTGTCTACCTGGACTGTCCCTGAGCAAAATCCACTAGAGGGTTGAGAAAGCTT 1305
Db 1187 GGATCTTCTGTCTACCTGGACTGTCCCTGAGCAAAATCCACTAGAGGGTTGAGAAAGCTT 1246
QY 1306 TGTGCGACCTGGTGTACAGCTCTGATCACTTTCTGAAATGCTTGTGAGCTTGCCTACA 1365
Db 1247 TGTGCGACCTGGTGTACAGCTCTGATCACTTTCTGAAATGCTTGTGAGCTTGCCTACA 1306
QY 1366 TGTGCGACCTGGTGTACAGCTCTGATCACTTTCTGAAATGCTTGTGAGCTTGCCTACA 1425
Db 1307 TGTGCGACCTGGTGTACAGCTCTGATCACTTTCTGAAATGCTTGTGAGCTTGCCTACA 1366
QY 1426 TGTGCGACGAGAAAGACTTTTGTGTCATGCTTGTGTCTAGAAACAGACTGGGAACT 1485
Db 1367 TGTGCGACGAGAAAGACTTTTGTGTCATGCTTGTGTCTAGAAACAGACTGGGAACT 1426
QY 1486 TATGTGAGCAGCAGATCCACAGAGTAAGAGGTAATGCTCTTCTTCTTCTGATCT 1545
Db 1427 TATGTGAGCAGCAGATCCACAGAGTAAGAGGTAATGCTCTTCTTCTTCTGATCT 1486
QY 1546 TCTGTGCTGGGAGACTTTCAGAGACTTGTGGCTGAGAGCCATTAACAGAGCAGCT 1605
Db 1487 TCTGTGCTGGGAGACTTTCAGAGACTTGTGGCTGAGAGCCATTAACAGAGCAGCT 1546
QY 1606 ATCAGTGAATGATCATTAACCTCCCTGTCACACTTGTGCCAATGGGGAATGATCT 1665
Db 1547 ATCAGTGAATGATCATTAACCTCCCTGTCACACTTGTGCCAATGGGGAATGATCT 1606
QY 1666 TTCCACCAAGAGCTCACAGACTTTTCCACAGAGATGC 1703
Db 1607 TTCCACCAAGAGCTCACAGACTTTTCCACAGAGATGC 1644

RESULT 7
AK074746 1979 bp mRNA linear PRI 03-SEP-2002
LOCUS AK074746
DEFINITION Homo sapiens cDNA FLJ90265 f1s, clone NTBRM4001325, weakly similar
TO CHONDROITIN 6-SULFOTRANSFERASE (EC 2.8.2.17).
ACCESSION AK074746

VERSION AK074746.1 GI:22760388
 KEYWORDS oligo capping; f1s (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 REFERENCE 1
 AUTHORS Isegai,T., Ota,T., Nishikawa,T., Hayashi,K., Otsuki,T.,
 Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S.,
 Kawai,Hio,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y.,
 Kojima,S., Nagahara,K., Masuno,Y., Ono,T., Okano,K., Yoshikawa,Y.,
 Aotsuka,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and
 Ninomiya,K.
 TITLE NEDO human cDNA sequencing project
 JOURNAL unpublished
 REFERENCE 2 (bases 1 to 1979)
 AUTHORS Isegai,T. and Otsuki,T.
 JOURNAL Direct Submission
 Submitted (25-MAR-2002) Takao Isegai, Helix Research Institute,
 Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:
 Research Association for Biotechnology; cDNA library construction:
 Institute of Medical Science, University of Tokyo, Laboratory of
 Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass
 sequencing and clone selection: Helix Research Institute (supported
 by Japan Key Technology Center etc.).
 FEATURES
 source
 1. 1979
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="NT2RM4001325"
 /cell_line="NT2"
 /cell_type="teratocarcinoma"
 /clone_id="NT2RM4"
 /note="cloning vector: PME18SFU3-mRNA from uninduced NT2
 neuronal precursor cells"
 69..1229
 /note="unamed protein product"
 /codon_start=1
 /protein_id="BAC11177.1"
 /db_xref="GI:22760389"
 /translation="MLPGRKKKLLFLYSOMAILALFEHMYSHNISLSMAKOPBRH
 VLVLSMRSGSFYGLGKHPDVFYLMPEPMHMYMPEKSFAMMLMAYVDLIRAY
 LCDMSYDATHNEPRPROSLFQENSRALCSAPACITIPDEITPRKCLLSQDP
 FEVEKACRSTSHVVLKVEFENLQSLPLKDPDLNHLIVLADPAVRSERERK
 GDLMDIRIVWGQHEOKLKEEDOPYVMQVQSQLELYTKTQSLPRALDORYLVRY
 EDLARAVVAQTSRMVEFVGLFELPHLOTVHNITRGKMGMDHAFHTVARADLANYSQAW
 RMSLPYKVSILKACGDAMNLLGVRHVSQEQBRNLLDLSTWYPEQTH"
 CDS
 BASE COUNT 451 a 542 c 474 g 512 t
 ORIGIN
 Query Match 75.7%; Score 1547; DB 9; Length 1979;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 346 GCCGCCCTGGACAGGTGGATGACCTTCACAGACAGCAGCGCTGGATGCTGCACATGGC 405
 DB 287 GCGCGCTGGACAGGTGGATGACCTTCACAGACAGCAGCGCTGGATGCTGCACATGGC 346
 QY 406 TGTGCGGGATGTGATACGGCGCTTCTTGTGCGACATGAGCGCTTTGATGGCTACAT 465
 DB 347 TGTGCGGGATGTGATACGGCGCTTCTTGTGCGACATGAGCGCTTTGATGGCTACAT 406
 QY 466 GGAACCTGGTCCCGGAGACAGTCCAGCTCTTTCAGTGGAGAAACAGCGCGGCTGTG 525
 DB 407 GGAACCTGGTCCCGGAGACAGTCCAGCTCTTTCAGTGGAGAAACAGCGCGGCTGTG 466
 QY 526 TTTCGACACCTGCTGTGATCATCATCCCAAGATGAATCATCCCGGGGCTCACTGAG 585
 DB 467 TTTCGACACCTGCTGTGATCATCATCCCAAGATGAATCATCCCGGGGCTCACTGAG 526
 QY 586 GCTCCTGTGACATCAACAGCCCTTGTAGGTGGTGAAGAGCGCTGCGCTCTACAGCA 645
 DB 527 GCTCCTGTGACATCAACAGCCCTTGTAGGTGGTGAAGAGCGCTGCGCTCTACAGCA 586
 QY 646 CGTGGTCTCAAGAGAGTGGCGCTTCTTCAACCTGCAGTCCCTTACCCGCTGTGAAGA 705
 DB 587 CGTGGTCTCAAGAGAGTGGCGCTTCTTCAACCTGCAGTCCCTTACCCGCTGTGAAGA 646
 QY 706 CCCCTCCCTCAACCTGCATATCGTGCACCTGCTGCGGAGCCCGGGGCGGTCCGTTCC 765
 DB 647 CCCCTCCCTCAACCTGCATATCGTGCACCTGCTGCGGAGCCCGGGGCGGTCCGTTCC 706
 QY 766 CCGAGAACGCACAAAGGAGATCTCATGATTTGACAGTGCATTTGTATGGGGGACAGATGA 825
 DB 707 CCGAGAACGCACAAAGGAGATCTCATGATTTGACAGTGCATTTGTATGGGGGACAGATGA 766
 QY 826 GCAAAACTCAAGAGAGAGACCAACCTACTATGTATGTGACAGTGCATTTGCCAAAGCCA 885
 DB 767 GCAAAACTCAAGAGAGAGACCAACCTACTATGTATGTGACAGTGCATTTGCCAAAGCCA 826
 QY 886 GCTGAGATCTACAAAGACATCATCAGTCTTCCCAAGGCGCTGCAGAACCTTCACTGCT 945
 DB 827 GCTGAGATCTACAAAGACATCATCAGTCTTCCCAAGGCGCTGCAGAACCTTCACTGCT 886
 QY 946 TGTGCGCTATGAGGACCTGCGTGCAGCCCTGTGGGCCAGACTTCCGATGTATGAAT 1005
 DB 887 TGTGCGCTATGAGGACCTGCGTGCAGCCCTGTGGGCCAGACTTCCGATGTATGAAT 946
 QY 1006 CGTGGATGTGAATTTCTGCCCCATCTTCAACACCTGGGTGCATTAACATCACCCAGGCAA 1065
 DB 947 CGTGGATGTGAATTTCTGCCCCATCTTCAACACCTGGGTGCATTAACATCACCCAGGCAA 1006
 QY 1066 GGGCATGGGTACCAAGCTTTCACACAAATGCGAGGATGCCCTTAATGTCTCCAGGC 1125
 DB 1007 GGGCATGGGTACCAAGCTTTCACACAAATGCGAGGATGCCCTTAATGTCTCCAGGC 1066
 QY 1126 TTGGCGCTGTCTTGGCCCTATGAAAAGTTTCTGACTGAGAAAAGCTGTGGCGATGC 1185
 DB 1067 TTGGCGCTGTCTTGGCCCTATGAAAAGTTTCTGACTGAGAAAAGCTGTGGCGATGC 1126
 QY 1186 CATGAATTTGCTGGGCTACCGCCACGTCAGATCTGAAGAACAAGACGAAACCTGTGCT 1245
 DB 1127 CATGAATTTGCTGGGCTACCGCCACGTCAGATCTGAAGAACAAGACGAAACCTGTGCT 1186
 QY 1246 GGATCTTCTGTCTACCTGAGCTGTCCCTGAGACCAATTCATAGAGGTTGAGAAGCTT 1305
 DB 1187 GGATCTTCTGTCTACCTGAGCTGTCCCTGAGACCAATTCATAGAGGTTGAGAAGCTT 1246
 QY 1306 TGTCTGACCTGAGTCAAGCTCAAGTCACTTTCTCTGAATCTTCTGAGCTTGCCTTACA 1365
 DB 1247 TGTCTGACCTGAGTCAAGCTCAAGTCACTTTCTCTGAATCTTCTGAGCTTGCCTTACA 1306
 QY 1366 TCTCTGACCTTAACTACATGCTGTGGGTATCAACACAGATGAGTGTGTCCACACG 1425
 DB 1307 TCTCTGACCTTAACTACATGCTGTGGGTATCAACACAGATGAGTGTGTGTCCACACG 1366

QY 1426 TCGTCAAGCAGAAAGACTTTTGTCTCCATGCTTGTCTTAGAAAAACAGACTGGGGAACCT 1485
Db 1367 TGTCTAAGCAGAAAGACTTTTGTCTCCATGCTTGTCTTAGAAAAACAGACTGGGGAACCT 1426
QY 1486 TATGTAGCAGCAGATCCACAGATGAAGAGGTTTGTCTTCTTCTTCTTGTATCT 1545
Db 1427 TATGTAGCAGCAGATCCACAGATGAAGAGGTTTGTCTTCTTCTTCTTGTATCT 1486
QY 1546 TCTGTCTGGGAGACTTCCAGAGACTTTGTGCGCTGGAGGCTATTAAAGCAGACAGT 1605
Db 1487 TCTGTCTGGGAGACTTCCAGAGACTTTGTGCGCTGGAGGCTATTAAAGCAGACAGT 1546
QY 1606 ATAGTGAATATATCATTAACCTCCCTGCTCCATCTTGGCCAAATGGGGAATGATCT 1665
Db 1547 ATAGTGAATATATCATTAACCTCCCTGCTCCATCTTGGCCAAATGGGGAATGATCT 1606
QY 1666 TTCACCAAGAGCTCAGACAGATTTTCCAGAGATGC 1703
Db 1607 TTCACCAAGAGCTCAGACAGATTTTCCAGAGATGC 1644

RESULT 8
AR203335 AR203335 2032 bp DNA linear PAT 20-JUN-2002

LOCUS AR203335
DEFINITION Sequence 2 from patent US 6365365.
ACCESSION AR203335
VERSION AR203335.1 GI:21499698
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2032)
AUTHORS Bistup,A., Rosen,S.D., Tangemann,K. and Hemmerich,S.
TITLE Method of determining whether an agent modulates glycosyl
sulfotransferase-3
JOURNAL Patent: US 6365365-A 2 02-APR-2002;
FEATURES
source location/Qualifiers
BASE COUNT 468 a 569 c 490 g 505 t

Query Match 75.7%; Score 1547; DB 6; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 106 CAAGTCTTCCACTTCCAGCAATGCTACTGCTAAAAAATGAGCTCTGCTTTCT 165
Db 160 CAAGTCTTCCACTTCCAGCAATGCTACTGCTAAAAAATGAGCTCTGCTTTCT 219
QY 166 GGTTCCTCCAGAGTGGCATCTTGGCTCTATTCTTCCAGATGTACAGCCACAAATCAGCTC 225
Db 220 GGTTCCTCCAGAGTGGCATCTTGGCTCTATTCTTCCAGATGTACAGCCACAAATCAGCTC 279
QY 226 CCGTGTATGAAGGACAGCCGAGGAGGAGTGCAGTGTCTGCTCTTCCGCGGCTC 285
Db 280 CCGTGTATGAAGGACAGCCGAGGAGGAGTGCAGTGTCTGCTCTTCCGCGGCTC 339
QY 286 TGGCTCTCTTTTGTGGGAGAGCTTTTGGGAGAGCCAGATGTTTCTACTGTATGA 345
Db 340 TGGCTCTCTTTTGTGGGAGAGCTTTTGGGAGAGCCAGATGTTTCTACTGTATGA 399
QY 346 GCCCGCTGGAGCAGTGTGATGACCTTCAAGCAGACACCGCTGTATGCTGACATGGC 405
Db 400 GCCCGCTGGAGCAGTGTGATGACCTTCAAGCAGACACCGCTGTATGCTGACATGGC 459
QY 406 TGTGGGAGATGTATACGGGCGCTCTTGTGTGCGACATGAGGCTTTTATGCTCATAT 465
Db 460 TGTGGGAGATGTATACGGGCGCTCTTGTGTGCGACATGAGGCTTTTATGCTCATAT 519
QY 466 GGAACCTGTGCTCCCGAGAGAGTCCAGCTCTTTCAGTGGGAGAAAGCGCGGCGCTGTG 525
Db 520 GGAACCTGTGCTCCCGAGAGAGTCCAGCTCTTTCAGTGGGAGAAAGCGCGGCGCTGTG 579

QY 526 TTCTGACCTGCTGTGACATCATCCACAAGATGAATATCCCGCGGCTCAGTGCAG 585
Db 580 TTCTGACCTGCTGTGACATCATCCACAAGATGAATATCCCGCGGCTCAGTGCAG 639
QY 586 GCTTCGTGTAGTCAACAGCCCTTTGAGTGTGTGAGAAAGGCTGCGCTCTACAGCCA 645
Db 640 GCTTCGTGTAGTCAACAGCCCTTTGAGTGTGTGAGAAAGGCTGCGCTCTACAGCCA 699
QY 646 CGTGTGCTCAAGAGAGTGGGCTTCTTCAACCTGAGTCCCTTACCGCTGTGAAGA 705
Db 700 CGTGTGCTCAAGAGAGTGGGCTTCTTCAACCTGAGTCCCTTACCGCTGTGAAGA 759
QY 706 CCCCTCCCTACCTGCAATATGCTGCACCTGTGCGGAGACCCCGGCGCTGTTCCGTT 765
Db 760 CCCCTCCCTACCTGCAATATGCTGCACCTGTGCGGAGACCCCGGCGCTGTTCCGTT 819
QY 766 CCGAGAACGACAAAGGAGATGTCATGATGACAGTGCATGTATGAGGAGAGATGA 825
Db 820 CCGAGAACGACAAAGGAGATGTCATGATGACAGTGCATGTATGAGGAGAGATGA 879
QY 826 GCAAAAACCTCAAGAGAGAGACCAACCTTACTATGTATGATCAGTCAATGCTGCCAAGCCA 885
Db 880 GCAAAAACCTCAAGAGAGAGACCAACCTTACTATGTATGATCAGTCAATGCTGCCAAGCCA 939
QY 886 GCTGAGATCTACAGACCATCCAGTCTTGTCCCAAGGCGCTGAGAGACGTAACCTGCT 945
Db 940 GCTGAGATCTACAGACCATCCAGTCTTGTCCCAAGGCGCTGAGAGACGTAACCTGCT 999
QY 946 TGTGGCTATGAGAGAGCTGCTGAGGCGCTGAGGCGCCAGACTTCCGAAATGATGAAT 1005
Db 1000 TGTGGCTATGAGAGAGCTGCTGAGGCGCTGAGGCGCCAGACTTCCGAAATGATGAAT 1059
QY 1006 CGTGGAGTTGGAATTTTGTGCCCCATCTTCCAGCTGAGTGCATATACATCAGCCGAGGCAA 1065
Db 1060 CGTGGAGTTGGAATTTTGTGCCCCATCTTCCAGCTGAGTGCATATACATCAGCCGAGGCAA 1119
QY 1066 GGGCATGGGTGACACAGCCTTCCACACAAATGCGAGGATGCCCTTAATGTCTCCAGGC 1125
Db 1120 GGGCATGGGTGACACAGCCTTCCACACAAATGCGAGGATGCCCTTAATGTCTCCAGGC 1179
QY 1126 TTGGCGCTGTCTTGGCCCTATGAAAAGTTTCTCACTTCAAAAAGCCGTGGGCGATGC 1185
Db 1180 TTGGCGCTGTCTTGGCCCTATGAAAAGTTTCTCACTTCAAAAAGCCGTGGGCGATGC 1239
QY 1186 CATGAATTTGCTGGGTACCGCCACGCTCAGATCTGAACAAGAACAGAAACCTGTGCT 1245
Db 1240 CATGAATTTGCTGGGTACCGCCACGCTCAGATCTGAACAAGAACAGAAACCTGTGCT 1299
QY 1246 GGAATCTTGTCTTACCTGACAGTCCCTGAGCAATTCACCTAAAGAGGTTGAGAGGCTT 1305
Db 1300 GGAATCTTGTCTTACCTGACAGTCCCTGAGCAATTCACCTAAAGAGGTTGAGAGGCTT 1359
QY 1306 TGTGCACTGCTGTGACAGCTGACAGTCTTCTGAAATGCTTGAAGCTTGGAGCTTCCATA 1365
Db 1360 TGTGCACTGCTGTGACAGCTGACAGTCTTCTGAAATGCTTGAAGCTTGGAGCTTCCATA 1419
QY 1366 TCTCTAGGCTTAACTACATGCTGTGAGTATCACAGTGTGATGTGTGTCCACAG 1425
Db 1420 TCTCTAGGCTTAACTACATGCTGTGAGTATCACAGTGTGATGTGTGTGTGTGTGTGTGT 1479
QY 1426 TGTCTAAGCAGAAAGACTTTTGTGCAATGCTTGTCTTAAGAAACAGACTGGGGAACCT 1485
Db 1480 TGTCTAAGCAGAAAGACTTTTGTGCAATGCTTGTCTTAAGAAACAGACTGGGGAACCT 1539
QY 1486 TATGTAGAGCAGATCCACAGCAGTGAAGAGGTTTGTCTTCTTCTTCTTCTTCTTCT 1545
Db 1540 TATGTAGAGCAGATCCACAGTGAAGAGGTTTGTCTTCTTCTTCTTCTTCTTCTTCTTCT 1599
QY 1546 TCTGTCTGTGGGAGACTTCAAGAGACTTGTGCGCTGAGAGCTTATTAAAGCAGACAGT 1605
Db 1600 TCTGTCTGTGGGAGACTTCAAGAGACTTGTGCGCTGAGAGCTTATTAAAGCAGACAGT 1659

Oy	1606	ATCAGTGAATTGATTCATTAACCTCCCGTCGCACACTCTGGCCCAATGGGAATGAGATCT	1605
Dd	1660	ATCATGTGAAATTGATTCATTAACCTCCCGTCGCACACTCTGGCCCAATGGGAATGAGATCT	1719
Oy	1666	TTCACCAAAGAGCTCACCAGCATTTTCCACAGAGATGC	1703
Dd	1720	TTCCACCAAGAAGCTCACACAGCATTTTCCACAGAGATGC	1757
RESULT 9	Bd134772		
LOCUS	Bd134772	2032 bp	DNA linear PAT 18-SEP-2002
DEFINITION	Glycosylsulfoltransferase-3.		
ACCESSION	Bd134772		
VERSION	Bd134772.1 GI:23229717		
KEYWORDS	JP 2002507409-A/1		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 2032) Bistrup,A., Rosen,S.D., Tangemann,K. and Hemmerich,S. Glycosylsulfoltransferase-3 Patent: JP 2002507409-A 1 12-MAR-2002;		
AUTHORS	THE REGENTS OF THE UNIVERSITY OF CALIFORNIA, SYNTAX INC		
TITLE	OS Homo sapiens (human)		
JOURNAL	PN JP 2002507409-A/1 PD 12-MAR-2002 PF 26-FEB-1999 JP 2000537979 PR 20-MAR-1998 US 09/045284,12-NOV-1998 US 09/190911 PI ANNETTE BISTRUP, STEVEN D ROSEN, KIRSTEN TANDEMANN, STEFAN PI HEMMERICH PC C12N9/10,A01K67/027,A61K45/00,A61K48/00,A61P29/00,A61P37/06, A61P43/00, PC A61P43/00, PC C12N1/15,C12N1/21,C12N5/10,C12N15/09,C12Q1/48,C12N5/00,C12N15/ PC 00 CC Glycosylsulfoltransferase-3 FC key Location/Qualifiers FT source 1..2032 Location/Qualifiers FT 1..2032 /organism='Homo sapiens (human)'. Location/Qualifiers 1..2032 /organism='Homo sapiens' /mol_type='genomic DNA' /db_xref='taxon:9606'		
FEATURES	source		
BASE COUNT	468 a 569 c 490 g 505 t		
ORIGIN			
Query Match	75.7%; Score 1547; DB 6; Length 2032;		
Best Local Similarity	99.9%; Pred. No. 0;		
Matches 1597; Conservative	0; Mismatches 1; Indels 0; Gaps 0;		
Oy	106	C AAGTGCTTCACATTCAGCACATACTGCTCTAAAAAATGAAGCTCGTGGTTTCT	165
Dd	160	C AAGTGCTTCACATTCAGCACATACTGCTCTAAAAAATGAAGCTCGTGGTTTCT	219
Oy	166	G GTTTCACAGATGGGCATCTGGGCGCTTTGTCACATGTAAGGCCACAACATCAGCTC	225
Dd	220	G GTTTCACAGATGGGCATCTGGGCGCTTTGTCACATGTAAGGCCACAACATCAGCTC	279
Oy	226	C CTGTCTATGAAAGGCACAGCCGAGCGCATGCAGCTGCTGTCTGCTTCCTGGCGCTC	285
Dd	280	C CTGTCTATGAAAGGCACAGCCGAGCGCATGCAGCTGCTGTCTGCTTCCTGGCGCTC	339
Oy	286	T TGCGTCTCTTTTGTGGGGAGCTTTTGGGACAGCACCAGATGTTTTCTACCTGATGGA	345
Dd	340	T TGCGTCTCTTTTGTGGGGAGCTTTTGGGACAGCACCAGATGTTTTCTACCTGATGGA	399
Oy	346	G GCCGCGCTGACACTGTGGATGACCTTCAACAGAGCACCCTGGATGCTGCACATGCG	405
Dd	400	G GCCGCGCTGACACTGTGGATGACCTTCAACAGAGCACCCTGGATGCTGCACATGCG	459
Oy	406	T TGTGGGAGATCTGATACGGGCGCTTCTTGTGTGGACATGAGCGCTTTTGATGCTTACAT	465

Db	460	TTGTGGGATCTGATACGGGCGCTTCTTGTGTGCGATAGAGCTTTTGATGCTTACAT	519
QY	466	GGAACTGGTCCCGGAGACAGTCCAGCCTCTTTCAGTGGGAAACAGCCGGGCGCTGTG	525
Db	520	GGAACTGGTCCCGGAGACAGTCCAGCCTCTTTCAGTGGGAAACAGCCGGGCGCTGTG	579
QY	526	TTTCTGACCTGCTGTGATCATATCCACAAATGAAATCATTCCTCCGGGCTCAGTCAG	585
Db	580	TTTCTGACCTGCTGTGATCATATCCACAAATGAAATCATTCCTCCGGGCTCAGTCAG	639
QY	586	GCTCTGTGACAGTCAACAGCCCTTGTAGGTGTGGAGAGGCGCTGCGCTCTACAGCA	645
Db	640	GCTCTGTGACAGTCAACAGCCCTTGTAGGTGTGGAGAGGCGCTGCGCTCTACAGCA	699
QY	646	GCTGTGCTCAAGGAGGTGCGCTTTCTCACTGCAAGTCCCTCTACCGCTGCTGAAGA	705
Db	700	GCTGTGCTCAAGGAGGTGCGCTTTCTCACTGCAAGTCCCTCTACCGCTGCTGAAGA	759
QY	706	CCCCCTCCAACTGCATATGCTGACCTGGTCCGGAGCCCGCGGCGGTGTCCGTTTC	765
Db	760	CCCCCTCCAACTGCATATGCTGACCTGGTCCGGAGCCCGCGGCGGTGTCCGTTTC	819
QY	766	CCGAGACGCGACAAAGGAGATCTCATGATTCACAGTCCGATTTGTATGGGCGACATGA	825
Db	820	CCGAGAGCGCACAAAGGAGATCTCATGATTCACAGTCCGATTTGTATGGGCGACATGA	879
QY	826	GCAAAAACTCAAAGAGGAGGACCAACCCACTATGTGATGACAGGTATGTGCCAAAGCCA	885
Db	880	GCAAAAACTCAAAGAGGAGGACCAACCCACTATGTGATGACAGGTATGTGCCAAAGCCA	939
QY	886	GCTGAGATCTACAAAGACCATCAAGTCTTTCGCCAAGGCCCTGCAGAAAGCTACCTGCT	945
Db	940	GCTGAGATCTACAAAGACCATCAAGTCTTTCGCCAAGGCCCTGCAGAAAGCTACCTGCT	999
QY	946	TTGTGGCATATGAGACCTGGGCTCGAGGCCCTGTGGCGCCAGACTTCCCGAATGATGAATT	100
Db	1000	TTGTGGCATATGAGACCTGGGCTCGAGGCCCTGTGGCGCCAGACTTCCCGAATGATGAATT	105
QY	1006	CGTGGATTTGGAATCTTGCCCATCTTGAGACTGGGTGATATACATCACCCGAGCA	106
Db	1060	CGTGGATTTGGAATCTTGCCCATCTTGAGACTGGGTGATATACATCACCCGAGCA	111
QY	1066	GGGCAATGGGTGACCAAGCTTTTCCACACAATGCGAGGATGGCCCTTAATGTCTCCAGGC	112
Db	1120	GGGCAATGGGTGACCAAGCTTTTCCACACAATGCGAGGATGGCCCTTAATGTCTCCAGGC	117
QY	1126	TTGGGCTGTGTTTGGCCTATGAAAAAGTTTCTGACATTCAAGAAAGCCTGTGGCATGC	118
Db	1180	TTGGGCTGTGTTTGGCCTATGAAAAAGTTTCTGACATTCAAGAAAGCCTGTGGCATGC	123
QY	1186	CATGAATTTGCTGGGCTACCGCACGTCAGATCTGACAAGAACAGAGAAACCTGTGTCT	124
Db	1240	CATGAATTTGCTGGGCTACCGCACGTCAGATCTGACAAGAACAGAGAAACCTGTGTCT	129
QY	1246	GGATCTTCTGTACTCTGACTGTGCCCTAGGAAATCCACTAAGAGGGTTGAGAAAGCTT	130
Db	1300	GGATCTTCTGTACTCTGACTGTGCCCTAGGAAATCCACTAAGAGGGTTGAGAAAGCTT	135
QY	1306	TGCTGCCACCTGGTGTACAGCTCAGTCACTTTCTGTGAATGCTTGTGAGCCTTGCCTTACA	136
Db	1360	TGCTGCCACCTGGTGTACAGCTCAGTCACTTTCTGTGAATGCTTGTGAGCCTTGCCTTACA	141
QY	1366	TCTTGAGCCTTAACTACATGTCTGTGGGTATCACACAGTAGTGTGATGTGTCCACAGC	142
Db	1420	TCTTGAGCCTTAACTACATGTCTGTGGGTATCACACAGTAGTGTGATGTGTCCACAGC	147
QY	1426	TGCTCAAGCAGAGAGACTTTTGTGTCAAGCTTGTGTAGAAAAACAGACTGGGAGAACT	148
Db	1480	TGCTCAAGCAGAGAGACTTTTGTGTCAAGCTTGTGTAGAAAAACAGACTGGGAGAACT	153
QY	1486	TATGTGACAGACATCCCAACAGTGAACAGGGATTTGCTCTTCTTCTTCTTGTGATCT	154

Db	1540	ATGTGACAGACACATCCACCAGTGAANAAGGGATATTGCCTTCCTCTTTCTTGTA	GACT	1599
Qy	1546	TCTGTCTGGGACAGACTTCACAGACTTTGTGGCCTTGAGGCCCTATTAAAGCACACAGT		1605
Db	1600	TCTGTCTGGGACAGACTTCACAGACTTTGTGGCCTTGAGGCCCTATTAAAGCACACAGT		1659
Qy	1606	ATCATGGTAATTGATATCCATAAAGCCCTCGTCACATCTTCCCAATGGGAATGAGACT		1665
Db	1660	ATCATGGTAATTGATATCCATAAAGCCCTCGTCACATCTTCCCAATGGGAATGAGACT		1719
Qy	1666	TTTCACCAAAGAGCTCACACGAGCTTTTCCACAGAGATGC		1703
Db	1720	TTTCACCAAAGAGCTCACACGAGCTTTTCCACAGAGATGC		1757
RESULT 10				
AFI31235		2032 bp	mRNA	linear PRI 10-AUG-1999
LOCUS				
DEFINITION	Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA,			
complete cds.				
AFI31235				
AFI31235.1	GI:4927113			
SOURCE	Homo sapiens (human)			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	Bistrup,A., Bhakta,S., Lee,J.K., Belov,Y.Y., Gunn,M.D., Zuo,F.R., Huang,C.C., Kannagi,R., Rosen,S.D. and Hemmerich,S.			
TITLE	Sulfotransferases of two specificities function in the reconstitution of high endothelial cell ligands for L-selectin			
JOURNAL	J. Cell Biol. 145 (4), 899-910 (1999)			
MEDLINE	99264336			
PUBMED	10330415			
REFERENCE	2 (bases 1 to 2032)			
AUTHORS	Bistrup,A., Tangemann,K., Bhakta,S., Lee,J.-K., Belov,Y.Y., Gunn,M.D., Zuo,F.-R., Huang,C.-C., Kannagi,R., Rosen,S.D. and Hemmerich,S.			
TITLE	Direct Submission			
JOURNAL	Submitted (25-FEB-1999) Respiratory Diseases, Roche Bioscience,			
FEATURES	3401 Hillview Avenue, Palo Alto, CA 94304, USA			
source	Location/Qualifiers			
CDS	1..2032			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/cell_type="high endothelial"			
	/tissue_type="tonsil"			
	182..1342			
	/function="biosynthesis of endothelial ligands for L-selectin."			
	/note="HEC-GlcNAc-6-S"			
	/codon_start=1			
	/product="N-acetylglucosamine 6-O-sulfotransferase"			
	/protein_id="AAD3015.1"			
	/db_xref="GI:4927114"			
	/translation="MLPKRKRLLLFVSOMALLFEFHMYSHNISSLSMAKPERMHLVLIWSRSGSFEGOLFQGHDPVEFLMEPAHWMTFKOSTAMLMHAAYDLRAVFALICDVSFDVAWMEPEPBRSSLFQWENSRALCSAPACIIIPODETIIPAHOILCSODPFEVEYKACRSYSHVLTKLEVPENSTSLXPLKDPSLNLHYLVADRPVFRSRERKGDMLIDRIYWGQHEOKLKKEDEDTYVMQYIOSQLTYKTGLPALDERLYLVYKEDLARAVADTSRYAEVGLGLEIFLPHLOTWHNTIRKGGMGHAFHTNARALNYSOAMRWMSLPYEKVSTIQACCDAMNLGLRYHRVRSOEORLNLDLSTWTVEQH"			
BASE COUNT	468 a 569 c 490 g 505 t			
ORIGIN				
Query Match	75.7%	Score 1547:	DB 9:	Length 2032:
Best Local Similarity	99.9%:	Pred. No. 0:		
Matches 1597:	Conservative 0:	Mismatches 1:	Indels 0:	Gaps 0:
QY	106	CAAGTCTTCACATTACGACACAATGCTACTGCTTAAAAAATGAAGCTCCCTGCTGTTCT		165

Db	160	CAGGCTCTCCACTTCACGACACATGCTACTGCTTAAAAAATGAAGCTCTGCTTTCT	219
Qy	166	GGTTTCCAGATGGCCATCTTGCGCTATTTCTTCCACATGTACAGCCACACATCAGCTC	225
Db	220	GGTTTCCAGATGGCCATCTTGCGCTATTTCTTCCACATGTACAGCCACACATCAGCTC	279
Qy	226	CCTGTCTATGAAGGACACAGCCCGAGCGCATGACGTGCTGGTTCGTCTTCTGCGCGTC	285
Db	280	CCTGTCTATGAAGGACACAGCCCGAGCGCATGACGTGCTGGTTCGTCTTCTGCGCGTC	339
Qy	286	TGGCTCTTCTTGTGTGGGGACACTTTTGTGGGACACCGAGATGTTTTCTTACCTGATGA	345
Db	340	TGGCTCTTCTTGTGTGGGGACACTTTTGTGGGACACCGAGATGTTTTCTTACCTGATGA	399
Qy	346	GCCCGCCTGGCACGTGTGATGACCTTCAAGCAGACACACCGCTGATGCTGCACATGGC	405
Db	400	GCCCGCCTGGCACGTGTGATGACCTTCAAGCAGACACACCGCTGATGCTGCACATGGC	459
Qy	406	TGTGGGGGATCTGATACGGGCGCTTCTTGTGGCACAATGAACGCTTTTATGCTTACAT	465
Db	460	TGTGGGGGATCTGATACGGGCGCTTCTTGTGGCACAATGAACGCTTTTGAAGCTTACAT	519
Qy	466	GGAACTGTGTCGCCGAGACAGTCCAGCCTCTTCAAGGGGAGAACAGCGCGGCGCTGTG	525
Db	520	GGAACTGTGTCGCCGAGACAGTCCAGCCTCTTCAAGGGGAGAACAGCGCGGCGCTGTG	579
Qy	526	TTCTGCACCTGCTGTGACATCATCCACAAAGATGAAATCATCCCGCGGCTCACTGACG	585
Db	580	TTCTGCACCTGCTGTGACATCATCCACAAAGATGAAATCATCCCGCGGCTCACTGACG	639
Qy	586	GCTCCTGTGTGACATCAACAGCCTTTTGAGTGGTGGAGAAAGCCTTCCGCTCTACAGCA	645
Db	640	GCTCCTGTGTGACATCAACAGCCTTTTGAGTGGTGGAGAAAGCCTTCCGCTCTACAGCA	699
Qy	646	CGTGTGCTCAAGGAGGTGCGCTTCTTCAACCTGTGAGTCCCTACCGGCTCTGAAGA	705
Db	700	CGTGTGCTCAAGGAGGTGCGCTTCTTCAACCTGTGAGTCCCTACCGGCTCTGAAGA	759
Qy	706	CCCTCCTCAACCTGCATATCTGTGACACCTGGTCCGGGACCCCGCGGCGTGTCCGTT	765
Db	760	CCCTCCTCAACCTGCATATCTGTGACACCTGGTCCGGGACCCCGCGGCGTGTCCGTT	819
Qy	766	CCGAAACGGCACAAAGGGAGATCTCTATGTATGACAGTGGCATTTGTGATGGGGACATGA	825
Db	820	CCGAAACGGCACAAAGGGAGATCTCTATGTATGACAGTGGCATTTGTGATGGGGACATGA	879
Qy	826	GC AAAAATCTCAAGAGGAGAGACCAACCTTACTATGTATGATGACAGTCTGTGCAAAAGCA	885
Db	880	GC AAAAATCTCAAGAGGAGAGACCAACCTTACTATGTATGATGACAGTCTGTGCAAAAGCA	939
Qy	886	GCTGAGATCTACAAGACATTCAGTCTTGCCCAAGGCCCTGTGACGAAAGCTTACTCTCT	945
Db	940	GCTGAGATCTACAAGACATTCAGTCTTGCCCAAGGCCCTGTGACGAAAGCTTACTCTCT	999
Qy	946	TGTGGGCTATAGAGACTGTGGCTCCAGGCCCTTGTGGCCCAAGACTTCCCAATGTATGAATT	1005
Db	1000	TGTGGGCTATAGAGACTGTGGCTCCAGGCCCTTGTGGCCCAAGACTTCCCAATGTATGAATT	1059
Qy	1006	CGTGGATTTGGAATCTTGCCCACTTTCAGACCTGGGGTGATATACATCACCCGAGGAA	1065
Db	1060	CGTGGATTTGGAATCTTGCCCACTTTCAGACCTGGGGTGATATACATCACCCGAGGAA	1119
Qy	1066	GGGCAATGGGTGACACGCTTTCACACAAATGCCAAGGATGCCCTTAATGTCTCCACAGC	1125
Db	1120	GGGCAATGGGTGACACGCTTTCACACAAATGCCAAGGATGCCCTTAATGTCTCCACAGC	1179
Qy	1126	TTGGCGGTGTCTTGGCCCTATGAAAAGTTTCTGCACTTACAGAAAGCCTGTGGCATGC	1185
Db	1180	TTGGCGGTGTCTTGGCCCTATGAAAAGTTTCTGCACTTACAGAAAGCCTGTGGCATGC	1239
Qy	1186	CATGAATTTGCTGGGCTACCGCGACGTCAAGATCTGAACAAAGACAGAAACCTGTCTCT	1245
Db	1240	CATGAATTTGCTGGGCTACCGCGACGTCAAGATCTGAACAAAGACAGAAACCTGTCTCT	1299

QY 1246 GGATCTTCTGCTACTGAGCTGTCCTGAGCAATGCAAGGGTTGAGAAGCTT 1305
 |||||||
 Db 1300 GGATCTTCTGCTACTGAGCTGTCCTGAGCAATGCAAGGGTTGAGAAGCTT 1359
 QY 1306 TGTGCGACCTGCTGTCAGCTGTCAGCTTCTGTAAGCTTCTGAGCTTGCCTACA 1365
 |||||||
 Db 1360 TGTGCGACCTGCTGTCAGCTGTCAGCTTCTGTAAGCTTCTGAGCTTGCCTACA 1419
 QY 1366 TCTGTGACCTTACTGATGCTGTCGTCATGTCAGCTGTCAGCTTGCCTACA 1425
 |||||||
 Db 1420 TCTGTGACCTTACTGATGCTGTCGTCATGTCAGCTGTCAGCTTGCCTACA 1479
 QY 1426 TGTGCAAGCAGAAAGAGCTTGTGTCATGCTTGTCTAGAAAACAGAGCTGGGAACT 1485
 |||||||
 Db 1480 TGTGCAAGCAGAAAGAGCTTGTGTCATGCTTGTCTAGAAAACAGAGCTGGGAACT 1539
 QY 1486 TATGTGAGCAGACATGTCAGCTGTCAGCTTGTCTAGAAAACAGAGCTGGGAACT 1545
 |||||||
 Db 1540 TATGTGAGCAGACATGTCAGCTGTCAGCTTGTCTAGAAAACAGAGCTGGGAACT 1599
 QY 1546 TCTGTGAGCAGACATGTCAGCTTGTGTCATGCTTGTCTAGAAAACAGAGCTGGGAACT 1605
 |||||||
 Db 1600 TCTGTGAGCAGACATGTCAGCTTGTGTCATGCTTGTCTAGAAAACAGAGCTGGGAACT 1659
 QY 1606 ATCAGTGAATTTGATCCATAAACCTCCCTGTCACATCTTGCCTAATGGGAAATGAGATCT 1665
 |||||||
 Db 1660 ATCAGTGAATTTGATCCATAAACCTCCCTGTCACATCTTGCCTAATGGGAAATGAGATCT 1719
 QY 1666 TTCACCAAGAGCTCACACGATTTTCCACAGAGATGC 1703
 |||||||
 Db 1720 TTCACCAAGAGCTCACACGATTTTCCACAGAGATGC 1757

RESULT 11
 AF149783 1333 bp mRNA linear PRI 02-JUL-2001

LOCUS Homo sapiens L-selectin ligand sulfotransferase mRNA, complete cds.
 DEFINITION AF149783
 ACCESSION AF149783
 VERSION AF149783.1 GI:13897503
 KEYWORDS

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1333)
 Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.

AUTHORS Yeh, J.-C., Hiraoka, N., Petryniak, B., Nakayama, J., Elles, L.G.,
 Rabuka, D., Hinds, J., Marth, J.D., Lowe, J.B., and Fukuda, M.
 Novel sulfated lymphocyte homing receptors and their control by a
 core1 extension beta 1,3-N-acetylgalactosaminyltransferase

JOURNAL Cell 105 (7), 957-969 (2001)
 MEDLINE 21332592
 PUBMED 11439191

REFERENCE 2 (bases 1 to 1333)
 Hiraoka, N. and Fukuda, M.
 Direct Submission
 Submitted (10-MAY-1999) Glycobiology, The Burnham Institute, 10901
 North Torrey Pines Road, La Jolla, CA 92037, USA

FEATURES
 source
 1. .1333
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 93. .1253
 /note="LST"
 /codon_start=1
 /product="L-selectin ligand sulfotransferase"
 /protein_id="AAK48417.1"
 /db_xref="GI:13897504"

CDS
 /translation="MLPKKKLLFLYSOMAILALFFHMYSHNISLSMAQPERKH
 VLVLSWRSSSEYVGLFGHDPVFLMEPWHMYWTFKOSTAMLMHAYDLIRAF
 LCDMSVDAYMEPGRSSILFQENSRALCSAPACDIIPDELIIIPRACHLLSDP
 FEVEKACRSYSHVLEKEVEFNLSLYPLLKPSLNLHIVHLVDRPRAVRSRERIK

1007 GTGGAGTGAATTCCTGCTCCCATCTTCAAGCTGGGTGATCAACATCACCGAGGCAAG 1066
 |||||||
 Db 912 GTGGAGTGAATTCCTGCTCCCATCTTCAAGCTGGGTGATCAACATCACCGAGGCAAG 971
 |||||||
 Db 972 GTGGAGTGAATTCCTGCTCCCATCTTCAAGCTGGGTGATCAACATCACCGAGGCAAG 1031

BASE COUNT 286 a 393 c 336 g 318 t
 ORIGIN

Query Match 61.8%; Score 1262; DB 9; Length 1333;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

107 AAGCTTCCACTTCAGCACAATGCTACTGCTTAAAAAATGAAGCTCTGCTTTTG 166
 |||||||
 Db 72 AAGCTTCCACTTCAGCACAATGCTACTGCTTAAAAAATGAAGCTCTGCTTTTG 131
 167 GTTCCCGATGGGATCTTGGCTATCTTCCAGATGTCAGGACACATGACGCTC 226
 |||||||
 Db 132 GTTCCCGATGGGATCTTGGCTATCTTCCAGATGTCAGGACACATGACGCTC 191
 227 CTGTCTATGAAGGACACAGCCGAGGCGATGACGCTGCTTCTCTCTGCGCTCT 286
 |||||||
 Db 192 CTGTCTATGAAGGACACAGCCGAGGCGATGACGCTGCTTCTCTCTGCGCTCT 251
 287 GGCCTCTTTTGTGGGACACTTTTGGGACAGACCCAGATGTTTCTACTGATGAG 346
 |||||||
 Db 252 GGCCTCTTTTGTGGGACACTTTTGGGACAGACCCAGATGTTTCTACTGATGAG 311
 347 CCGGCTGGCAGCGTGGATGACCTTCAAGCAGACACCGCTGGATGCTGACATGCT 406
 |||||||
 Db 312 CCGGCTGGCAGCGTGGATGACCTTCAAGCAGACACCGCTGGATGCTGACATGCT 371
 407 GTGGGGATCTGATACGGGCGCTCTTCTGTGCGACATGAGCGTCTTGTATGCTTAC 466
 |||||||
 Db 372 GTGGGGATCTGATACGGGCGCTCTTGTGCGACATGAGCGTCTTGTATGCTTAC 431
 467 GAACCTGTCCTCCCGAGACATGCAAGCTCTTTCAGTGGGAGAACACCGGCGCTGCT 526
 |||||||
 Db 432 GAACCTGTCCTCCCGAGACATGCAAGCTCTTTCAGTGGGAGAACACCGGCGCTGCT 491
 527 TCTGACCTGCTGTCATCATCCCAAGATGAATCAATCCCGGCGCTGCTGAGG 586
 |||||||
 Db 492 TCTGACCTGCTGTCATCATCCCAAGATGAATCAATCCCGGCGCTGCTGAGG 551
 587 CTCCTGTGACATCAACACCCCTTGTAGGTGTGAGGAGAGGCGCTGCTCTACAGCCAC 646
 |||||||
 Db 552 CTCCTGTGACATCAACACCCCTTGTAGGTGTGAGGAGAGGCGCTGCTCTACAGCCAC 611
 647 GTGGTCTCAAGAGAGTCCGCTTCTTCAACTGCAAGTCCCTCTACCCGCTGTAAGAC 706
 |||||||
 Db 612 GTGGTCTCAAGAGAGTCCGCTTCTTCAACTGCAAGTCCCTCTACCCGCTGTAAGAC 671
 707 CCGTCCCAACCTGCAATGATGTGACACTGTGCGGGAGCCCGGCGCGCTGCTGCTCC 766
 |||||||
 Db 672 CCGTCCCAACCTGCAATGATGTGACACTGTGCGGGAGCCCGGCGCGCTGCTGCTCC 731
 767 CGAAGACGACAAAGAGAGATCTCATGTTTACAGTGCATTTGATGGGACAGATGAG 826
 |||||||
 Db 732 CGAAGACGACAAAGAGAGATCTCATGTTTACAGTGCATTTGATGGGACAGATGAG 791
 827 CAAAAACTCAAGAGAGAGACCAACCTTACTATGTGATGACAGTCACTGCCAAGCCAG 886
 |||||||
 Db 792 CAAAAACTCAAGAGAGAGACCAACCTTACTATGTGATGACAGTCACTGCCAAGCCAG 851
 887 CTGAGATCTCAAGAGACATGCTGCTTGGCCAAAGGCGCTGCAAGAACGCTTACTGCTT 946
 |||||||
 Db 852 CTGAGATCTCAAGAGACATGCTGCTTGGCCAAAGGCGCTGCAAGAACGCTTACTGCTT 911
 947 GTGGCTATGAGAGCTGGCTGAGCCCTGTGGCCAGACTTCCGAAATGATGAATTC 1006
 |||||||
 Db 912 GTGGCTATGAGAGCTGGCTGAGCCCTGTGGCCAGACTTCCGAAATGATGAATTC 971
 1007 GTGGAGTGAATTCCTGCTCCCATCTTCAAGCTGGGTGATCAACATCACCGAGGCAAG 1066
 |||||||
 Db 972 GTGGAGTGAATTCCTGCTCCCATCTTCAAGCTGGGTGATCAACATCACCGAGGCAAG 1031

OY		1067	GGCAATGGGTGACACAGCGCTTTCACACAATAATGCCAGGGAATGCCCTTAATGTTCCAGGCT	1128
Dd		1092	GGCATGGGTGCACACAGCTTTCACACAATAATGCCAGGGAATGCCCTTAATGTTCCAGGCT	1091
OY		1127	TGGCGCTGGTCTTTTGCCCTATGAAAGAAGTTTCTGACTTCAGAAAAGCCTGTGGCGATGCC	1186
Dd		1092	TGGCGCTGGTCTTTTGCCCTATGAAAGAAGTTTCTGACTTCAGAAAAGCCTGTGGCGATGCC	1151
OY		1187	ATGAATTGTCTGGGGCTACCGCCAGCTGAGATCTGAACAAGACAGAAAACTGTGCTG	1246
Dd		1152	ATGAATTGTCTGGGGCTACCGCCAGCTGAGATCTGAACAAGACAGAAAACTGTGCTG	1211
OY		1247	GATCTTCTGTACTGACTGATGTGCTCCCTAGCAAAATCCATAAGAGGGTTGGAAGGCTT	1306
Dd		1212	GATCTTCTGTACTGACTGATGTGCTCCCTAGCAAAATCCATAAGAGGGTTGGAAGGCTT	1271
OY		1307	GCTGCCACCTGGTGTCAAGCTCTCACTTCTCTGTAATGCTTCTGAGCCCTTGCTACAT	1366
Dd		1272	GCTGCCACCTGGTGTCAAGCTCTCACTTCTCTGTAATGCTTCTGAGCCCTTGCTACAT	1331
OY		1367 CT 1368		
Dd		1332 CT 1333		
RESULT 12				
LOCUS	ARJ10485	2065 bp	DNA	linear PAT 12-JUN-2003
DEFINITION	Sequence 26 from patent US 6558935.			
ACCESSION	ARJ10485			
VERSION	ARJ10485.1 GI:31703448			
KEYWORDS	.			
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	Unclassified.			
AUTHORS	1 (bases 1 to 2065) Tang,Y.T., Corley,N.C., Guegler,K.J., Baughn,M.R., Ial,P., Yue,H., Hillman,J.L. and Azimzal,Y. Human transferase proteins Patent: US 6558935-A 26 06-MAY-2003; location/qualifiers 1..2065 /organism="unknown"			
TITLE				
JOURNAL				
FEATURES				
source				
BASE COUNT	466 a 575 c 491 g 533 t			
ORIGIN				
Query Match	56.4%; Score 1153; DB 6; Length 2065;			
Best Local Similarity	99.6%; Pred. NO. 0;			
Matches 1593; Conservative	0; Mismatches 4; Indels 2; Gaps 2;			
OY		106	CNAGGTCTTCACATTCAGCACAAATGCTACTGCTAATAAATAATGAAGCTCGTGTTCCT	165
Dd		152	CNAGGTCTTCACATTCAGCACAAATGCTACTGCTAATAAATAATGAAGCTCGTGTTCCT	211
OY		166	GGTTTCCCAGATGCGCATCTTGCTCTAATTCCTCAATGTACAGCCAAACATCACTC	225
Dd		212	GGTTTCCCAGATGCGCATCTTGCTCTAATTCCTCAATGTACAGCCAAACATCACTC	271
OY		226	CCCTGCTATGAAGGACACAGCCCGAGGCGATGACAGTCTGCTGCTTCCTCTGCGGCTC	285
Dd		272	CCCTGCTATGAAGGACACAGCCCGAGGCGATGACAGTCTGCTGCTTCCTCTGCGGCTC	331
OY		286	TGCGCTCTCTTTTGTGGGGGAGCTTTTGTGGGACACACCAGATGTTTCTAACGTGAGA	345
Dd		332	TGCGCTCTCTTTTGTGGGGGAGCTTTTGTGGGACACACCAGATGTTTCTAACGTGAGA	391
OY		346	GCCCCGCTGGCAGCTGTGAGTACCTTCAAGACAGACCGCCCTGATGCTGCACATGGC	405
Dd		392	GCCCCGCTGGCAGCTGTGAGTACCTTCAAGACAGACCGCCCTGATGCTGCACATGGC	451
OY		406	TGTGCGGGGANTGATTAGGGGCGCTTCTTGTGGCAATGAGCGCTTTGTATGCTTACAT	465

Db	452	TGTGGGGGATCTGATACGGGGCCGTCTTCTTGTCGACATAGAGCCCTTTGATGCGCTACAT	511
QY	466	GGAACTCTGTCCTCCCGAGACAGTCCAGCCTCTTTAGTGGGAAACAGCCGGGCCCTGTG	525
Db	512	GGAACTCTGTCCTCCCGAGACAGTCCAGCCTCTTTAGTGGGAAACAGCCGGGCCCTGTG	571
QY	526	TTCTGCACCTGCTGTGACATCATCCACAAGATGAATTCATCCCGGGGCTCACTGGAC	585
Db	572	TTCTGCACCTGCTGTGACATCATCCACAAGATGAATTCATCCCGGGGCTCACTGGAC	630
QY	586	GCTCCTGTGCACTACACAGCCCTTTG - AGTGTGTGGAGAGGCCCTGCGCTCTACAGCC	644
Db	631	GCTCCTGTGCACTACACAGCCCTTTGAAAGTTGTGGAGAAAAGCCGCGCTCTACAGCC	690
QY	645	ACGTGTCTCAAGAGAGGTGGCGCTTCAACCTGCAGTCCCTCAACCGGCTGCGAAG	704
Db	691	ACGTGTCTCTAAGAGAGGTGGCGCTTCAACCTGCAGTCCCTCAACCGGCTGCGAAG	750
QY	705	ACCCCTCCCTCAACCTGCATATCGTGACCTGTGTCGGGAGCCCGCGGGCCGTTCGGTT	764
Db	751	ACCCCTCCCTCAACCTGCATATCGTGACCTGTGTCGGGAGCCCGCGGGCCGTTCGGTT	810
QY	765	CCCGAAGACGACAAAGGAGATCATCATGTAACAGTCGATTCGATTTGATGGGCGACATG	824
Db	811	CCCGAAGACGACAAAGGAGATCATCATGTAACAGTCGATTTGATGGGCGACATG	870
QY	825	AGCAAAAACCTCAAGAGAGAGACACAAACCTCTATGTATGATGAGGTCACTGTCCAAACC	884
Db	871	AGCAAAAACCTCAAGAGAGAGACACAAACCTCTATGTATGATGAGGTCACTGTCCAAACC	930
QY	885	AGCTGGAGATCTACAAGACACATCCAGTCTCTTGCCCAAGGCCCTGAGAGAACGCTACCTGC	944
Db	931	AGCTGGAGATCTACAAGACACATCCAGTCTCTTGCCCAAGGCCCTGAGAGAACGCTACCTGC	990
QY	945	TTGTGCGGTATGAGAGACCTGTGCTGAGCCCGCTGTGGGCCAGACTCCCGAATGTATGAT	1004
Db	991	TTGTGCGGTATGAGAGACCTGTGCTGAGCCCGCTGTGGGCCAGACTCCCGAATGTATGAT	1050
QY	1005	TCGTGGGATTTGGAAATCTTGGCCCATCTTCAGACTGGGTGATTAACATCACCCGAGCA	1066
Db	1051	TCGTGGGATTTGGAAATCTTGGCCCATCTTCAGACTGGGTGATTAACATCACCCGAGCA	1110
QY	1065	AGGCATGGGAGACACAGCTTCTCCACACAAATGCCAGGATGGCCCTTAATGTCGCCAGG	1124
Db	1111	AGGCATGGGAGAGACACAGCTTCTCCACACAAATGCCAGGATGGCCCTTAATGTCGCCAGG	1170
QY	1125	CTTGGCGCTGTCCTTGGCCCTATGAAAAGTTTCTCGACTTCAGAAAACCTGTGGCGATG	1184
Db	1171	CTTGGCGCGTGTCTTGGCCCTATGAAAAGTTTCTCGACTTCAGAAAACCTGTGGCGATG	1230
QY	1185	CCATGAATTTCTGTGGGCTACGCCACGTCACATCTGAAACAAGAACAGAAAACCTGTGTC	1244
Db	1231	CCATGAATTTCTGTGGGCTACGCCACGTCACATCTGAAACAAGAACAGAAAACCTGTGTC	1290
QY	1245	TGATATCTTCTCTACCTGAGACTGCTCCCTGAGCAAAATCCACTAAGAGGTTGAGAAAGCT	1304
Db	1291	TGATATCTTCTCTACCTGAGACTGCTCCCTGAGCAAAATCCACTAAGAGGTTGAGAAAGCT	1350
QY	1305	TTTGGTGGCACCTGGGTGTAGCCTCAGTACTTTCTCTAATGCTTCTAGCCTTGCCCTAC	1366
Db	1351	TTTGGTGGCACCTGGGTGTAGCCTCAGTACTTTCTCTAATGCTTCTAGCCTTGCCCTAC	1410
QY	1365	ATCTCTGAGCCTTAACATCATATGTCGTGGGTATCACACTGAGTGAATTGTCTCCACAC	1424
Db	1411	ATCTCTGAGCCTTAACATCATATGTCGTGGGTATCACACTGAGTGAATTGTCTCCACAC	1470
QY	1425	GTGCTCAAGCAGAAAGACTTTTGTCTCATGCTTGTCTAGAAAACAGACTGGGGAAACC	1484
Db	1471	GTGCTCAAGCAGAAAGACTTTTGTGTCTCATGCTTGTCTAGAAAACAGACTGGGGAAACC	1530
QY	1485	TTATGTAGACAGACATATCCACACAGTAAACAGGGATATGCTCTTTCTTTCTTGATC	1544
Db	1531	TTATGTAGACAGACATATCCACACAGTAAACAGGGATATGCTCTTTCTTTCTTGATC	1590

QY 1545 TTCTGTCGTGGGAGAGCTTCAGAGACTTTGGGCTGGAGGCTATTAAAGCAGACAG 1604
 DB 1591 TTCTGTCGTGGGAGAGCTTCAGAGACTTTGGGCTGGAGGCTATTAAAGCAGACAG 1650
 QY 1605 TATGAGTGGAGTATATCCATTAACCTCCCTGTCACATCTTGCCCAATGGGAAATGATC 1664
 DB 1651 TATGAGTGGAGTATATCCATTAACCTCCCTGTCACATCTTGCCCAATGGGAAATGATC 1710
 QY 1665 TTTCACCAAGAGCTCCACGACATTTTCCACAGAGATGC 1703
 DB 1711 TTTCACCAAGAGCTCCACGACATTTTCCACAGAGATGC 1749
 RESULT 13
 BD124832
 LOCUS BD124832 877 bp DNA linear PAT 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD124832
 VERSION BD124832.1 GI:23219777
 KEYWORDS JP 2002017375-A/263
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 877)
 Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
 Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
 Koga,H.
 Primer for synthesizing full-length cDNA and use thereof
 Patent: JP 2002017375-A 263 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/263
 PD 22-JAN-2002
 PE 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TETSUJI OTSUKI,HISASHI KOGA
 PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
 10,
 PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 Location/Qualifiers
 FT source 1..877
 Location/Qualifiers
 1..877
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 183 a 266 c 221 g 203 t 4 others
 ORIGIN
 Query Match 34.0%; Score 695; DB 6; Length 877;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 106 CAAGGTCTTCACCTTCAGACAAATGCTACTGCTTAAATAATGAAGCTCTGCTTCT 165
 DB 47 CAAGGTCTTCACCTTCAGACAAATGCTACTGCTTAAATAATGAAGCTCTGCTTCT 106
 QY 166 GGTTCACAGATGGCCATCTGGCTATTCTTCCACATGTACAGCCACAATCATAGCTC 225
 DB 107 GGTTCACAGATGGCCATCTGGCTATTCTTCCACATGTACAGCCACAATCATAGCTC 166
 QY 226 CCTGTATGAAAGCAGACGCCGAGCGCATGACAGTGTGCTGCTTCTCTGCGCTC 285
 DB 167 CCTGTATGAAAGCAGACGCCGAGCGCATGACAGTGTGCTGCTTCTCTGCGCTC 226

QY 286 TGCGTCTCTTTTGTGGGGAGAGCTTTTGGGAGACACACAGATGTTTCTACATGATGA 345
 DB 227 TGCGTCTCTTTTGTGGGGAGAGCTTTTGGGAGACACACAGATGTTTCTACATGATGA 286
 QY 346 GCCCGCTGGCAGGTGTGATGATCCTTCACAGACAGACCGCTGTGATGTGCACATGGC 405
 DB 287 GCCCGCTGGCAGGTGTGATGATCCTTCACAGACAGACCGCTGTGATGTGCACATGGC 346
 QY 406 TGTGGGGATGTGAATGAGGGGCGTCTTGTGGAGACATGAGCGCTTGTGATGCTACAT 465
 DB 347 TGTGGGGATGTGAATGAGGGGCGTCTTGTGGAGACATGAGCGCTTGTGATGCTACAT 406
 QY 466 GGAACCTGTGTCCTGGAGACAGTCCAGGCTCTTCAGTGGAGAAACACCGGGCTGTG 525
 DB 407 GGAACCTGTGTCCTGGAGACAGTCCAGGCTCTTCAGTGGAGAAACACCGGGCTGTG 466
 QY 526 TTTCGACCTGCTGTCGATCATATCCACAAGATGAATATCCCGGGGCTCATCTGAG 585
 DB 467 TTTCGACCTGCTGTCGATCATATCCACAAGATGAATATCCCGGGGCTCATCTGAG 526
 QY 586 GCTCCTGTGAGTCAACAGACCGCTTTGAGGTGTGAGAGAGCGCTGCTCTACAGCA 645
 DB 527 GCTCCTGTGAGTCAACAGACCGCTTTGAGGTGTGAGAGAGCGCTGCTCTACAGCA 586
 QY 646 CGTGCTGTCAAGAGAGTGGCGCTTCTTCAACCTGACAGTCCCTTACCGGCTGTGAAGA 705
 DB 587 CGTGCTGTCAAGAGAGTGGCGCTTCTTCAACCTGACAGTCCCTTACCGGCTGTGAAGA 646
 QY 647 CCCCTCCCTCAACCTGCATATGTCGACCTGTCGGGAGACCCCGGGCGCTGCTTC 706
 DB 706 CCCCTCCCTCAACCTGCATATGTCGACCTGTCGGGAGACCCCGGGCGCTGCTTC 705
 QY 766 CCGAGAACGCAAGAGAGATGTCATGATTTGACA 800
 DB 707 CCGAGAACGCAAGAGAGATGTCATGATTTGACA 741
 RESULT 14
 BD126950
 LOCUS BD126950 877 bp DNA linear PAT 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD126950
 VERSION BD126950.1 GI:23221895
 KEYWORDS JP 2002017375-A/2381.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 877)
 Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
 Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
 Koga,H.
 Primer for synthesizing full-length cDNA and use thereof
 Patent: JP 2002017375-A 2381 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/2381
 PD 22-JAN-2002
 PE 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TETSUJI OTSUKI,HISASHI KOGA
 PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
 10,
 PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 Location/Qualifiers
 FT source 1..877
 Location/Qualifiers
 1..877
 /organism="Homo sapiens (human)".
 FEATURES
 Location/Qualifiers

	9	306	15.0	2988	21	AAC76156	Human OEPX ORF1711
	10	114	5.6	389	24	ABK54794	Human colon cancer
C	11	105	5.1	548	22	AAK92588	Human CDNA 3'-end
	12	59	2.9	1647	24	AAD24670	Human drug metabol
	13	59	2.9	1694	22	AAAD02701	Human glycosyl sul
	14	59	2.9	2044	22	AAD02699	Human glycosyl sul
	15	59	2.9	2170	22	AAD02698	Human glycosyl sul
	16	59	2.9	2544	24	ABN89536	Human corneal N-ac
	17	59	2.9	48436	24	ABN89503	Human corneal N-ac
	18	59	2.9	160552	22	AAD02697	Human glycosyl sul
	19	27	1.3	1926	20	AAZ20793	Mouse glycosyl sul
	20	23	1.1	1937	24	AAI65948	Murine intestinal-1
	21	23	1.1	1989	22	AAD02696	Mouse glycosyl sul
	22	21	1.0	10004	22	ABA14483	Human nervous syst
	23	21	1.0	10246	22	ABA14485	Human nervous syst
	24	20	1.0	354	22	AAF65813	Novel human polynn
	25	20	1.0	370	22	AAF65813	Novel human polynn
	26	20	1.0	381	22	AAF67389	Novel human polynn
C	27	20	1.0	388	25	ABX45934	Bovine EST associa
C	28	20	1.0	3763	19	AAV58194	Human myosin L-chain
C	29	20	1.0	3763	23	AAV58192	DNA encoding novel
C	30	20	1.0	5814	24	ACC46455	Human ovary specif
	31	20	1.0	6147	25	ACC46455	Human dhtp proteol
C	32	20	1.0	9516	22	AAK73555	Human lmmune/haema
C	33	20	1.0	9516	22	AAK81394	Human lmmune/haema
C	34	20	1.0	13192	22	AAK73556	Human lmmune/haema
C	35	20	1.0	13192	22	AAK81396	Human lmmune/haema
C	36	20	1.0	35959	22	AAK78275	Human lmmune/haema
C	37	20	1.0	180557	24	ABN85450	Human BAC clone RP
C	38	20	1.0	349980	22	ABL86431	Pycococcus abyssi
C	39	19	0.9	341	24	ABL85863	Human ovarian canc
C	40	19	0.9	341	24	ABL85910	Human ovarian canc
C	41	19	0.9	348	22	AAF67078	Novel human polynn
	42	19	0.9	349	22	AAI65385	Human prostate CDN
	43	19	0.9	349	24	ABI95356	Human prostate-spe
	44	19	0.9	349	25	ACA59793	Prostate cancer th
	45	19	0.9	353	22	AAF66848	Novel human polynn

ALIGNMENTS

RESULT 1	
AAK94229	AAK94229 standard; cDNA; 1979 BP.
ID	
XX AAK94229;	
AC	
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human full-length cDNA, SEQ ID NO: 2816.
XX	
KW	Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX	
OS	Homo sapiens.
XX	
PN	EPI130094-A2.
XX	
PD	05-SEP-2001.
XX	
PF	07-JUL-2000; 2000EP-0114089.
XX	
PR	08-JUL-1999; 99JP-0194486.
PR	11-JAN-2000; 2000JP-0118774.
PR	02-MAY-2000; 2000JP-0183765.
XX	
PA	(HELI-) HELIX RES INST.
PI	Ota T., Nishikawa T., Isogai T., Hayashi K., Ishii S., Kawai Y.
PI	Matsumatsu A., Sugiyama T., Nagai K., Kojima S., Otsuki T., Koga H.
OR	WPI; 2001-524255/58.
OR	P-PSDB; AAM93309.

XX 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 XX
 PS Claim 8; SEQ ID NO 2816; 1380bp + sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5' - and 3' - ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a full length
 CC human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 XX
 SO Sequence 1979 BP; 451 A; 542 C; 474 G; 512 T; 0 other;

Query Match 75.7%; Score 1547; DB 22; Length 1979;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 106 CAAGGCTTCCACTTACGACAAATGCTACTGCTTAAATAAATGAGCTCTGTTTCT 165
 DB 47 CAAGGCTTCCACTTACGACAAATGCTACTGCTTAAATAAATGAGCTCTGTTTCT 106
 QY 166 GATTTCACAGATGGCATCTTGCTCTATCTCCAGATGACGACACAAATGAGCTC 225
 DB 107 GATTTCACAGATGGCATCTTGCTCTATCTCCAGATGACGACACAAATGAGCTC 166
 QY 226 CTTGTATGAAAGGACACAGCCGAGCGATGACGATGCTGCTCTCTCTGAGGCTC 285
 DB 167 CTTGTATGAAAGGACACAGCCGAGCGATGACGATGCTGCTCTCTCTGAGGCTC 226
 QY 286 TGCTCTCTCTTCTTGTGGGAGCTTTTGGGACGACCCAGATGTTTCTACTGATGGA 345
 DB 227 TGCTCTCTCTTCTTGTGGGAGCTTTTGGGACGACCCAGATGTTTCTACTGATGGA 286
 QY 346 GCCGCGCTGACAGTGTGATGACCTTCAAGCAGAGACACCGCTGATGCTGACATGGC 405
 DB 287 GCCGCGCTGACAGTGTGATGACCTTCAAGCAGAGACACCGCTGATGCTGACATGGC 346
 QY 406 TGTGCGGAGTCTGATACGGGCGCTCTTGTGCGACATGAGCGCTTTGATGCTACAT 465
 DB 347 TGTGCGGAGTCTGATACGGGCGCTCTTGTGCGACATGAGCGCTTTGATGCTACAT 406
 QY 466 GGAACTGTGCTCCCGGAGACAGTCCAGCCCTTTCAGTGGGAAAGACCGCGGCTGTG 525
 DB 407 GGAACTGTGCTCCCGGAGACAGTCCAGCCCTTTCAGTGGGAAAGACCGCGGCTGTG 466
 QY 526 TTCTGACCTGCTGTGATCATATCCCAAGATGAATCATCCCGGCGCTACTGAC 585
 DB 467 TTCTGACCTGCTGTGATCATATCCCAAGATGAATCATCCCGGCGCTACTGAC 526
 QY 586 GCTCCTGTGATGACAGACCCCTTGAAGTGTGGAAGAGCGCTGCGCTCTACAGCA 645
 DB 527 GCTCCTGTGATGACAGACCCCTTGAAGTGTGGAAGAGCGCTGCGCTCTACAGCA 586
 QY 646 CGTGTGCTCAAGGAGTGGCTTTCACCTGAGTCCCTTACCCGCTGTGAAGA 705
 DB 587 CGTGTGCTCAAGGAGTGGCTTTCACCTGAGTCCCTTACCCGCTGTGAAGA 646
 QY 706 CCCCTCCCTCAACCTCATATGCTGACAGTGTGCGGAGCCCGGCGGTTCCTGTT 765
 DB 647 CCCCTCCCTCAACCTCATATGCTGACAGTGTGCGGAGCCCGGCGGTTCCTGTT 706
 QY 766 CCGAGAACGACAAAGGAGATCTCATGATGACAGTCCGATGTTGATGGGAGCATGA 825
 DB 707 CCGAGAACGACAAAGGAGATCTCATGATGACAGTCCGATGTTGATGGGAGCATGA 766

QY 826 GCAAAATCTCAAGAGAGGAGACCAACCTACTTGTGATGACAGTACTGCCAAAGCA 885
 DB 767 GCAAAATCTCAAGAGAGGAGACCAACCTACTTGTGATGACAGTACTGCCAAAGCA 826
 QY 886 GCTGAGATCTACAGACCATTCAGTCTTGCCCAAGGCGCTCAGAAACGTAACGCT 945
 DB 827 GCTGAGATCTACAGACCATTCAGTCTTGCCCAAGGCGCTCAGAAACGTAACGCT 886
 QY 946 TGTGCGCTATGAGAGACTGCTGAGACCCCTGTGGCCAGACTTCCGATGATGAT 1005
 DB 887 TGTGCGCTATGAGAGACTGCTGAGACCCCTGTGGCCAGACTTCCGATGATGAT 946
 QY 1006 CTTGGATGGAATTTCTTGCCCATTTTCAGACCTGAGTATACATCAACCCGAGCA 1065
 DB 947 CTTGGATGGAATTTCTTGCCCATTTTCAGACCTGAGTATACATCAACCCGAGCA 1006
 QY 1066 GGGCATGGGAGACAGCTTTCACCAAAATGCAAGGAGTCCCTTAATGTCGCCAGC 1125
 DB 1007 GGGCATGGGAGACAGCTTTCACCAAAATGCAAGGAGTCCCTTAATGTCGCCAGC 1066
 QY 1126 TTGGCGCTGCTCTTGGCCCTATGAAAAGTTTCTGACTTCAGAAAAGCTGTGGCATGC 1185
 DB 1067 TTGGCGCTGCTCTTGGCCCTATGAAAAGTTTCTGACTTCAGAAAAGCTGTGGCATGC 1126
 QY 1186 CATGAATTTGCTGGCTACCGCCAGCTCAGATGTGAACAAGAACAGAACTGTGCT 1245
 DB 1127 CATGAATTTGCTGGCTACCGCCAGCTCAGATGTGAACAAGAACAGAACTGTGCT 1186
 QY 1246 GGAATTTGCTACCGGAGCTGCTCCCTGAGCAATCCACTAGAGGTTGAGAAAGCTT 1305
 DB 1187 GGAATTTGCTACCGGAGCTGCTCCCTGAGCAATCCACTAGAGGTTGAGAAAGCTT 1246
 QY 1306 TGTGACACCTGCTGTCAAGCTCAGTCACTTCTCTGAATGCTTCTGAGCTTGCCTACA 1365
 DB 1247 TGTGACACCTGCTGTCAAGCTCAGTCACTTCTCTGAATGCTTCTGAGCTTGCCTACA 1306
 QY 1366 TCTCTAGCTTACTACTACATGTCTGGGATATACACTGATGATGCTGTCCACAG 1425
 DB 1307 TCTCTAGCTTACTACTACATGTCTGGGATATACACTGATGATGCTGTCCACAG 1366
 QY 1426 TGTCTAAGCAGAAAGACTTTTGTGTCATGCTGTGTAGAAACAGACTGGGAACT 1485
 DB 1367 TGTCTAAGCAGAAAGACTTTTGTGTCATGCTGTGTAGAAACAGACTGGGAACT 1426
 QY 1486 TATGTAGCAGACATCCACAGAGTAAACAGGATTTGCTCTCTCTTCTTCTGATCT 1545
 DB 1427 TATGTAGCAGACATCCACAGAGTAAACAGGATTTGCTCTCTCTTCTTCTGATCT 1486
 QY 1546 TCTGTGCTGGGAGACTTTCAGAGACTTTGGGCTGGAGGCTTATTAACACAGACAT 1605
 DB 1487 TCTGTGCTGGGAGACTTTCAGAGACTTTGGGCTGGAGGCTTATTAACACAGACAT 1546
 QY 1606 ATCAGTGAATGATCATTAACCTCCCTGTCACATCTTGGCCAAATGGGGAATGATCT 1665
 DB 1547 ATCAGTGAATGATCATTAACCTCCCTGTCACATCTTGGCCAAATGGGGAATGATCT 1606
 QY 1666 TTACCAAAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
 DB 1607 TTACCAAAAGAGCTCACAGCATTTTCCACAGAGATGC 1644

RESULT 2

AAZ20792
 ID AAZ20792 standard; DNA; 2032 BP.

AAZ20792;
 AC

XX 08-DEC-1999 (first entry)

XX Human glycosyl sulfotransferase-3 coding sequence.

DE Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
 XX selectin binding interaction; inflammation; lymphocyte homing; human;
 KW

KW secondary lymph organ; ss.
 XX Homo sapiens.
 XX MO9949018-A1.
 XX 30-SEP-1999.
 XX
 PE 26-FEB-1999; 99MO-US04316.
 XX
 PR 20-MAR-1998; 98US-0045284.
 PR 12-NOV-1998; 98US-0190911.
 XX
 PA (REGC) UNIV CALIFORNIA.
 PA (SYNT) SYNTX USA INC.
 XX
 PI Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
 DR WPI: 1999-580442/49.
 DR P-PSDB; AAY3918.
 XX
 PT Human and murine glycosyl sulfotransferase 3 and related
 XX polynucleotides
 PS
 XX Claim 4; Fig 1; 59pp; English.
 CC This sequence encodes the human glycosyl sulfotransferase-3 (GST-3) of
 CC the invention. The nucleic acid sequences, probes and primers derived
 CC from these, proteins and antibodies are useful in detecting homologues.
 CC The sequences, antibodies and methods are useful in the diagnosis and
 CC treatment of diseases associated with selectin binding interactions,
 CC including conditions associated with or resulting from the homing of
 CC leukocytes to sites of inflammation and the normal homing of lymphocytes
 CC to secondary lymph organs.
 CC
 XX Sequence 2032 BP; 468 A; 569 C; 490 G; 505 T; 0 other:
 XX
 Query Match 75.7%; Score 1547; DB 20; Length 2032;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 586 GCTCTGTGACATGACAGCCCTTTGAGGTGGTGGAGAGGCTGCCGCTCTACAGCA 645
 DB 640 GCTCTGTGACATGACAGCCCTTTGAGGTGGTGGAGAGGCTGCCGCTCTACAGCA 699
 QY 646 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTCAGTCCCTTACCCGCTGTGAAGA 705
 DB 700 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTCAGTCCCTTACCCGCTGTGAAGA 759
 QY 706 CCCCTCCCTCAACCTGATATGCGACACCTGTGTCGGGAGCCCGGCGCTCCGTTCC 765
 DB 760 CCCCTCCCTCAACCTGATATGCGACACCTGTGTCGGGAGCCCGGCGCTCCGTTCC 819
 QY 766 CCGAGAAAGCACAAGAGAGATCTCATGATGACAGTGCATTTGTGATGGGAGCAGATGA 825
 DB 820 CCGAGAAAGCACAAGAGAGATCTCATGATGACAGTGCATTTGTGATGGGAGCAGATGA 879
 QY 826 GCAAAAACCTCAAGAGAGAGACCAACCTTACTATGTGATGAGGTCAATGCCAAGCCA 885
 DB 880 GCAGAAACTCAAGAGAGAGACCAACCTTACTATGTGATGAGGTCAATGCCAAGCCA 939
 QY 886 GCTGAGATCTACAAAGACATCCAGTCCCTTCCCAAGGCCCTGAGAGAGCTACCTGCT 945
 DB 940 GCTGAGATCTACAAAGACATCCAGTCCCTTCCCAAGGCCCTGAGAGAGCTACCTGCT 999
 QY 946 TGTGCGTATGAGAGACCTGCTCGAGCCCTGTGGCCAGACTTCCGAAATGTATGAATT 1005
 DB 1000 TGTGCGTATGAGAGACCTGCTCGAGCCCTGTGGCCAGACTTCCGAAATGTATGAATT 1059
 QY 1006 CGTGGATTTGAATTTCTTGGCCCATCTTCAGACCTGGGTGATTAATCACCAGGCA 1065
 DB 1060 CGTGGATTTGAATTTCTTGGCCCATCTTCAGACCTGGGTGATTAATCACCAGGCA 1119
 QY 1066 GGGCATGGGTACCAAGCTTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGC 1125
 DB 1120 GGGCATGGGTACCAAGCTTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGC 1179
 QY 1126 TTGGCGGTGCTTTTGGCCCATGAAAGTTTCTGACTTCAGAAAGCCTGTGGCATGC 1185
 DB 1180 TTGGCGGTGCTTTTGGCCCATGAAAGTTTCTGACTTCAGAAAGCCTGTGGCATGC 1239
 QY 1186 CATGAATTTGCTGGGCTACCCGACAGTCAATCTGAACAGACAGAAACCTGTGCT 1245
 DB 1240 CATGAATTTGCTGGGCTACCCGACAGTCAATCTGAACAGACAGAAACCTGTGCT 1299
 QY 1246 GGATCTTCTGCTACCTGGAGATGCCGAGACAAATCCATAGAGGTTGAGAAGGCTT 1305
 DB 1300 GGATCTTCTGCTACCTGGAGATGCCGAGACAAATCCATAGAGGTTGAGAAGGCTT 1359
 QY 1306 TGTGCGACCTGTGTCAGCTCAGTCACTTCTGTAATGCTTCTGAGCCTTGCCTACA 1365
 DB 1360 TGTGCGACCTGTGTCAGCTCAGTCACTTCTGTAATGCTTCTGAGCCTTGCCTACA 1419
 QY 1366 TCTGTGACCTTAACTACATGTCTGTGGATTCACACTGAGTGTGAGTGTGTCCACAG 1425
 DB 1420 TCTGTGACCTTAACTACATGTCTGTGGATTCACACTGAGTGTGAGTGTGTCCACAG 1479
 QY 1426 TGTCTAAGCANAAGAGCTTTGTGTGCTGCTTGTGTATAAAAACAGACGCGGAACCT 1485
 DB 1480 TGTCTAAGCANAAGAGCTTTGTGTGCTGCTTGTGTATAAAAACAGACGCGGAACCT 1539
 QY 1486 TATGTAGACAGACATCCACAGTGAAGAGGGATATGCTCTTCTTTTCTTGATCT 1545
 DB 1540 TATGTAGACAGACATCCACAGTGAAGAGGGATATGCTCTTCTTTTCTTGATCT 1599
 QY 1546 TCTGTGTGGGACAGCTTACAGACTTGTGGCTGTGAGGCTTATTAAGACGACAGCT 1605
 DB 1600 TCTGTGTGGGACAGCTTACAGACTTGTGGCTGTGAGGCTTATTAAGACGACAGCT 1659
 QY 1606 ATCAGTGAATTTGATCCATAAATCCCTGTGTCACATCTTGGCCAAATGGGGAATGATCT 1665
 DB 1660 ATCAGTGAATTTGATCCATAAATCCCTGTGTCACATCTTGGCCAAATGGGGAATGATCT 1719
 QY 1666 TTCACAAAGAGCTCACAGCATTTTCCACAGAGATGC 1703

DB 1720 TTCACCAAGAGCTACACAGCATTTTCACAGAGATGC 1757

RESULT 3

AA16947 standard; cDNA: 1333 BP.

AA16947:

12-MAR-2002 (first entry)

Human L-selectin sulfotransferase-2 (LST-2) cDNA.

Human: beta1,3GNT; beta1,3-N-acetylglucosaminyltransferase; MECA-79; ss; L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes; ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus; allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2; delayed-type hypersensitivity reaction; hyperplastic thymus; anti-leuc; anti-inflammatory; antipsoriatic; antidiabetic; dermatological; antiallergic.

Homo sapiens.

Key Location/Qualifiers

CDS 111..1253

/tag= a

/product= "Human LST-2"

WO200185177-A1..

15-NOV-2001.

10-MAY-2001; 2001MO-US15452.

11-MAY-2000; 2000US-0569320.

(BURN-) BURNHAM INST.

Fukuda M, Yeh J, Hirakawa N;

WPI; 2002-075226/10.

P-PSDB; AAU11274.

New enzyme, useful for modifying acceptor molecule, comprises an isolated L-selectin sulfotransferase-2 that directs expression of L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or intestinal GlcNAc 6-sulfotransferase

Claim 19; Fig 4; 98pp; English.

The present invention provides a method of modifying an acceptor molecule by contacting the acceptor with an isolated

beta1,3-N-acetylglucosaminyltransferase (beta1,3GNT) or an active fragment, where beta1,3GNT directs expression of a MECA-79 antigen. The

invention also provides a method of treating or preventing an L-selectin-mediated condition by reducing the expression or activity of a

beta1,3GNT that directs expression of a MECA-79 antigen. This can be done by administering to the subject an oligosaccharide L-selectin antagonist

that inhibits binding of L-selectin to a MECA-79 antigen, for example by administering antibody material that specifically binds beta1,3GNT,

and/or a beta1,3GNT antisense nucleic acid molecule. L-selectin

sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.

Alternatively, the expression or activity of LST-2 or its active

fragment can be reduced in combination with reducing the expression or

activity of beta1,3GNT. The method is useful for treating L-selectin

mediated conditions such as Crohn's disease and ulcerative colitis,

inflammatory disorders of the skin such as allergic contact dermatitis,

psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type

hypersensitivity reactions, diabetes and hyperplastic thymus. This

sequence represents cDNA encoding human LST-2.

Sequence 1333 BP; 286 A; 393 C; 336 G; 318 T; 0 other;

Query Match 61.8%; Score 1262; DB 24; Length 1333;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	107	AAGTCTTCACCTTACAGACCAATGCTACCTTAAATAAGAGCTCTGCTTTCTG	166
DB	72	AAGTCTTCACCTTACAGACCAATGCTACCTTAAATAAGAGCTCTGCTTTCTG	131
OY	167	GTTCACCAAGAGCTACACAGCATTTTCACAGAGATGC 1757	226
DB	132	GTTCACCAAGAGCTACACAGCATTTTCACAGAGATGC 1757	191
OY	227	CTGTATATGAGGACAGAGCCGAGCCGATGACAGCTGCTGCTTCTTCTGCGCTC	286
DB	192	CTGTATATGAGGACAGAGCCGAGCCGATGACAGCTGCTGCTTCTTCTGCGCTC	251
OY	287	GGCTCTTCTTTTGTGGGAGAGCTTTTGGGAGAGCCAGATTTTCTACCTGATGAG	346
DB	252	GGCTCTTCTTTTGTGGGAGAGCTTTTGGGAGAGCCAGATTTTCTACCTGATGAG	311
OY	347	CCGCGCTGGACAGCTGTGATGACCTTCAAGACAGAGCAGCCGCTGATGCTGACATGGCT	406
DB	312	CCGCGCTGGACAGCTGTGATGACCTTCAAGACAGAGCAGCCGCTGATGCTGACATGGCT	371
OY	407	GTGCGGAGTCTGATACGAGGCGCTCTTCTTGTGAGATGAGCGCTTTGATGCTTACATG	466
DB	372	GTGCGGAGTCTGATACGAGGCGCTCTTCTTGTGAGATGAGCGCTTTGATGCTTACATG	431
OY	467	GAACCTGTGTCCTGGGAGACAGATCCACCTTCTTCAATGAGGAGAACAGCCGCGCTGTGT	526
DB	432	GAACCTGTGTCCTGGGAGACAGATCCACCTTCTTCAATGAGGAGAACAGCCGCGCTGTGT	491
OY	527	TCTGACAGCTGCTGTACATCATCCCAAGATGATCATCCCGGCGCTCACTGACAG	586
DB	492	TCTGACAGCTGCTGTACATCATCCCAAGATGATCATCCCGGCGCTCACTGACAG	551
OY	587	CTCTGTGACATGACAGCCCTTTGAGTGTGAGAGAGGCTGCGCTCTTACAGCCAC	646
DB	552	CTCTGTGACATGACAGCCCTTTGAGTGTGAGAGAGGCTGCGCTCTTACAGCCAC	611
OY	647	GTGCTGCTCAAGAGGAGGCTGCTTCAACCTGACCTGCTTACCCGCTGCGGAGAG	706
DB	612	GTGCTGCTCAAGAGGAGGCTGCTTCAACCTGACCTGCTTACCCGCTGCGGAGAG	671
OY	707	CCCTCCCTCAACCTGATATGTCGACCTGTGCGGAGCCCGGCGCTGTCCGCTTCC	766
DB	672	CCCTCCCTCAACCTGATATGTCGACCTGTGCGGAGCCCGGCGCTGTCCGCTTCC	731
OY	767	CGAGAACGCAAAAGGAGATCTCATGATTTGACAGTCGATTTGATGGGCGACATGAG	826
DB	732	CGAGAACGCAAAAGGAGATCTCATGATTTGACAGTCGATTTGATGGGCGACATGAG	791
OY	827	CAAAATCTAAGAGAGAGGACCAACCTTATGATGAGGAGCTTCTGCAAAAGCCAG	886
DB	792	CAAAATCTAAGAGAGAGGACCAACCTTATGATGAGGAGCTTCTGCAAAAGCCAG	851
OY	887	CTGAGATCTACAAAGACATCCAGCTTGTGCGGAGGCGCTGAGAGAGCTTCTGCTT	946
DB	852	CTGAGATCTACAAAGACATCCAGCTTGTGCGGAGGCGCTGAGAGAGCTTCTGCTT	911
OY	947	GTGCGTATGAGAGCTGCTGAGGCGCTTGTGCGGAGGCGCTTCCGAGATGTGATTTT	1006
DB	912	GTGCGTATGAGAGCTGCTGAGGCGCTTGTGCGGAGGCGCTTCCGAGATGTGATTTT	971
OY	1007	GTGCGATTTGGAATTTCTGCGGAGCTTGTGAGGCTGCTGATGATGATGATGATGATG	1066
DB	972	GTGCGATTTGGAATTTCTGCGGAGCTTGTGAGGCTGCTGATGATGATGATGATGATG	1031
OY	1067	GGCATGGTGAGACAGCTTTCACACAAATGCGAGGAGTCCCTTATGTCTCCAGGCT	1126
DB	1032	GGCATGGTGAGACAGCTTTCACACAAATGCGAGGAGTCCCTTATGTCTCCAGGCT	1091


```

Db      991  TTTGTGGCTATGAGGAGCTGGCTGGAGCCCTGTGGCCAGACTTCCGATATATGAAT 1050
QY      1005  TCGTGGGATTTGAAATCTTGCCCATCTTTCAGACTGGGTCATTAATCAACCCGAGCA 1064
Db      1051  TCGTGGGATTTGAAATCTTGCCCATCTTTCAGACTGGGTCATTAATCAACCCGAGCA 1110
QY      1065  AGGGCATGGGTGACCCAGCTTTCCACAAAGTCCAGAGGATGCCCTTAATGTCTCCAGG 1124
Db      1111  AGGGCATGGGTGACCCAGCTTTCCACAAAGTCCAGAGGATGCCCTTAATGTCTCCAGG 1170
QY      1125  CTTGGCGCTGGTCTTGGCCCTATGAAAGGTTCTGACTCTGAAAGGCTGGGGATG 1184
Db      1171  CTTGGCGCTGGTCTTGGCCCTATGAAAGGTTCTGACTCTGAAAGGCTGGGGATG 1230
QY      1185  CCATGAATTTGCTGGGCTACCCGACGTGATCTGACAAAGAGAGAAACCTGTGC 1244
Db      1231  CCATGAATTTGCTGGGCTACCCGACGTGATCTGACAAAGAGAGAAACCTGTGC 1290
QY      1245  TGGATCTTCTGTCTACTGAGCTGCTCCCTGAGCAATTCACATAAGGGTTGAAAGCT 1304
Db      1291  TGGATCTTCTGTCTACTGAGCTGCTCCCTGAGCAATTCACATAAGGGTTGAAAGCT 1350
QY      1305  TTGCTGCCACCTGGTGTACAGCTCAGTCACTTCTCTGAAATGCTTCTGAGCCCTGCTAC 1364
Db      1351  TTGCTGCCACCTGGTGTACAGCTCAGTCACTTCTCTGAAATGCTTCTGAGCCCTGCTAC 1410
QY      1365  ATCTCTGAGCCTTAACATCATGTCTGGGTATCACACTGATGATGATGTCACAC 1424
Db      1411  ATCTCTGAGCCTTAACATCATGTCTGGGTATCACACTGATGATGATGTCACAC 1470
QY      1425  GTGCTCAAGAGAGAGACTTTGTGTCATGCTGTGTCTAGAAACAGACCTGGGAAC 1484
Db      1471  GTGCTCAAGAGAGAGACTTTGTGTCATGCTGTGTCTAGAAACAGACCTGGGAAC 1530
QY      1485  TATGTGAGACAGACATCCACAGTAAGAAACAGAGATTTGCTCTTCTTCTTGATC 1544
Db      1531  TATGTGAGACAGACATCCACAGTAAGAAACAGAGATTTGCTCTTCTTCTTGATC 1590
QY      1545  TTCCCTCTGGGAGACTTCAGAGACTTTGTGCTGGAGGCTTATTAAGACAGACAG 1604
Db      1591  TTCCCTCTGGGAGACTTCAGAGACTTTGTGCTGGAGGCTTATTAAGACAGACAG 1650
QY      1605  TATCACTGGAATTCATCAATAACCTCCCTGTCACATCTTGGCCCAATGGGAATGATC 1664
Db      1651  TATCACTGGAATTCATCAATAACCTCCCTGTCACATCTTGGCCCAATGGGAATGATC 1710
QY      1665  TTTTACCAAAAGAGCTCAGCAGATTTTCCACAGAGATGC 1703
Db      1711  TTTTACCAAAAGAGCTCAGCAGATTTTCCACAGAGATGC 1749

RESULT 5
AAK91803
ID AAK91803 standard; cDNA; 877 BP.
XX
AC AAK91803;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human cDNA 5'-end sequence, SEQ ID NO: 263.
XX
KM Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
OS Homo sapiens.
XX
PN EP130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.

```

```

PR      11-JAN-2000; 2000JP-0118774.
PR      02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
DR WPI; 2001-524255/58.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
PS
XX Claim 2; SEQ ID NO 263; 1380bp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is the nucleotide
CC sequence of the 5'-end of a cDNA provided in the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;
XX

Query Match      34.0%; Score 695; DB 22; Length 877;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      106  CAAGGTCTTCCACTGAGCACAATGCTACTGCTATAAAATAGCTCTGCTTTCT 165
Db      47  CAAGGTCTTCCACTGAGCACAATGCTACTGCTATAAAATAGCTCTGCTTTCT 106
QY      166  GGTTTCCAGATGGCCATCTTGCTCTATTCTTCCACATGTACAGCACAATCAGCTC 225
Db      107  GGTTTCCAGATGGCCATCTTGCTCTATTCTTCCACATGTACAGCACAATCAGCTC 166
QY      226  CCTGTATTAAGAGCAGACGCCAGGCGATGACAGGCTGCTGCTCTGCTGGGCGTC 285
Db      167  CCTGTATTAAGAGCAGACGCCAGGCGATGACAGGCTGCTGCTCTGCTGGGCGTC 226
QY      286  TGGCTCTCTTTTGTGGGCGAGCTTTTGGGCGACACCAGATGTTTCTACCTGATGA 345
Db      227  TGGCTCTCTTTTGTGGGCGAGCTTTTGGGCGACACCAGATGTTTCTACCTGATGA 286
QY      346  GCCCGCTGGCAGAGTGTGATGTACCTTCAAGCAGAGACCGCCTGGATGTGCACATGGC 405
Db      287  GCCCGCTGGCAGAGTGTGATGTACCTTCAAGCAGAGACCGCCTGGATGTGCACATGGC 346
QY      406  TGTGGGAGATCTGATACGGGCGCTCTTGTGCGACATAGAGGCTCTTGTATGCTACAT 465
Db      347  TGTGGGAGATCTGATACGGGCGCTCTTGTGCGACATAGAGGCTCTTGTATGCTACAT 406
QY      466  GGAACCTGTGCTCCCGAGAGACAGTCCAGCTCTTTCAGTGGGAGAACAGCGGCGCTGTG 525
Db      407  GGAACCTGTGCTCCCGAGAGACAGTCCAGCTCTTTCAGTGGGAGAACAGCGGCGCTGTG 466
QY      526  TTTGTGACCTGCTGCTGACATCATCTCCACAAAGTAAATCATATCCCGGGGCTCACGCA 585
Db      467  TTTGTGACCTGCTGCTGACATCATCTCCACAAAGTAAATCATATCCCGGGGCTCACGCA 526
QY      586  GCTCTGTGAGTCAAGCAGACGCCCTTGAAGGTGGAGAGGCTGCGCTCTACAGCCA 645
Db      527  GCTCTGTGAGTCAAGCAGACGCCCTTGAAGGTGGAGAGGCTGCGCTCTACAGCCA 586
QY      646  CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGCAGTCCCTCTACCCGCTGTGAAGA 705
Db      587  CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGCAGTCCCTCTACCCGCTGTGAAGA 646

```

QY 706 CCCCTCCCTCAACCTGATATGTCACCTGTCGGGAGCCCGGGCGCTGTTCCGTTTC 765
 DB 647 CCCCTCCCTCAACCTGATATGTCACCTGTCGGGAGCCCGGGCGCTGTTCCGTTTC 706
 QY 766 CCGAGAACGACAAAGGAGATCTCATGATTGACA 800
 DB 707 CCGAGAACGACAAAGGAGATCTCATGATTGACA 741

RESULT 6
 ID AAK93921 standard; cDNA: 877 BP.
 AC AAK93921;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human cDNA clone representative sequence, SEQ ID NO: 2381.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EPI130094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PE 07-JUL-2000; 2000EP-0114089.
 XX
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0189765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI; 2001-524255/58.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 XX
 PS Example 11; SEQ ID NO 2381; 1380bp + sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence was used in
 CC representative sequence from a human clone which was used in
 CC homology searches to identify the clone.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 CC
 SQ Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;

Query Match 34.0%; Score 695; DB 22; Length 877;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 CAAGCTTCCACTTACGACACATGCTACTGCTTAAATAAATGAAGCTCTGCTGTTTCT 165
 DB 47 CAAGCTTCCACTTACGACACATGCTACTGCTTAAATAAATGAAGCTCTGCTGTTTCT 106
 QY 166 GGTTCCTCCAGATGGCATCTGGCTCTATCTTCCACATGTACACCCACATCAGCTC 225
 DB 107 GGTTCCTCCAGATGGCATCTGGCTCTATCTTCCACATGTACACCCACATCAGCTC 166

QY 226 CCTGCTATGAAGGACACAGCCCGAGCGCATGCGATGCTGTTCTGCTTCCTGCGGCTC 285
 DB 167 CCTGCTATGAAGGACACAGCCCGAGCGCATGCGATGCTGTTCTGCTTCCTGCGGCTC 226
 QY 286 TGGCTCTTCTTTTGTGGGGCAGCTTTTGGGAGCACCAGATGTTTCTACATGATGGA 345
 DB 227 TGGCTCTTCTTTTGTGGGGCAGCTTTTGGGAGCACCAGATGTTTCTACATGATGGA 286
 QY 346 GCCCGCTTGGCAGCTGTGATGACCTTTCAAGCAGACCGGCTTGATGTCACATGGC 405
 DB 287 GCCCGCTTGGCAGCTGTGATGACCTTTCAAGCAGACCGGCTTGATGTCACATGGC 346
 QY 406 TGTGGGATGTGATGATGAGGCGCTTCTTGTGGGATGAGCGCTTTGATGCTACAT 465
 DB 347 TGTGGGATGTGATGATGAGGCGCTTCTTGTGGGATGAGCGCTTTGATGCTACAT 406
 QY 466 GGAACCTGTGTCCTCCGAGACAGTCCAGGCTTTTCAGTGGGAGAACACCGGCGCTGTG 525
 DB 407 GGAACCTGTGTCCTCCGAGACAGTCCAGGCTTTTCAGTGGGAGAACACCGGCGCTGTG 466
 QY 526 TTCTGACCTGCTGCTGTGACATCATCCACAGATGAATATCCCGGCGCTCACTGAG 585
 DB 467 TTCTGACCTGCTGCTGTGACATCATCCACAGATGAATATCCCGGCGCTCACTGAG 526
 QY 586 GCTCCTGTGCAATCAAGAGCCCTTTGAGGTGGTGGAGAGGCGCTGCTCTACAGCA 645
 DB 527 GCTCCTGTGCAATCAAGAGCCCTTTGAGGTGGTGGAGAGGCGCTGCTCTACAGCA 586
 QY 646 GGTGTGCTCAAGAGGTGCGCTTCTTCAACCTGAGTCCCTTACCGCTGCTGAAGA 705
 DB 587 GGTGTGCTCAAGAGGTGCGCTTCTTCAACCTGAGTCCCTTACCGCTGCTGAAGA 646
 QY 706 CCCCTCCCTCAACCTGATATGTCACCTGTCGGGAGCCCGGGCGCTGTTCCGTTTC 765
 DB 647 CCCCTCCCTCAACCTGATATGTCACCTGTCGGGAGCCCGGGCGCTGTTCCGTTTC 706
 QY 766 CCGAGAACGACAAAGGAGATCTCATGATTGACA 800
 DB 707 CCGAGAACGACAAAGGAGATCTCATGATTGACA 741

RESULT 7
 ID ABV89280 standard; cDNA: 505 BP.
 AC ABV89280;
 XX
 DT 13-DEC-2002 (first entry)
 XX
 DE Human colon cancer related cDNA SEQ ID NO 2595.
 XX
 KW Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
 KW ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200258534-A2.
 XX
 PD 01-AUG-2002.
 XX
 PE 19-NOV-2001; 2001WO-US43704.
 PF
 PR 20-NOV-2000; 2000US-252222P.
 PR 06-FEB-2001; 2001US-267011P.
 PR 28-MAR-2001; 2001US-279670P.
 PR 10-JUL-2001; 2001US-304037P.
 XX
 XX (CORI-) CORIXA CORP.
 PI Stolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GE;
 XX
 DR WPI; 2002-608400/65.

XX New isolated tumor colon polynucleotide and polypeptide, useful for the
 PT diagnosis, prevention and/or treatment of cancer, in particular colon
 PT cancer
 XX
 PS Claim 1: SEQ ID NO 2595; 266pp + Sequence Listing: English.
 CC
 CC The invention relates to a human colon tumour expressed polynucleotide
 CC (I) encoding a polypeptide (II), ABP67991-ABP67996) comprising: (i) any of
 CC 2600 fully defined nucleotide sequences (ABV6669-ABV8289); (ii)
 CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
 CC sequences that hybridize to (i) under moderately stringent conditions;
 CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
 CC degenerate variants of (i). The compositions and methods of the present
 CC invention are useful for the diagnosis, prevention and/or treatment of
 CC cancer, particularly colon cancer. (i) can be used in gene therapy and
 CC (ii) and (iii) are useful in pharmaceutical compositions such as vaccines.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pcl_sequences.
 CC
 SO Sequence 505 BP; 120 A; 125 C; 119 G; 141 T; 0 other;
 XX
 Query Match 19.7%; Score 403; DB 24; Length 505;
 Best Local Similarity 100.0%; Pred. No. 2.6e-189;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1301 GGCTTGGTCCACCTGCTGTCAGCTCAGTCACTTCTGTAATGCTTGTGAGCCTTGC 1360
 Db 1 GGGTTTGGTCCACCTGCTGTCAGCTCAGTCACTTCTGTAATGCTTGTGAGCCTTGC 60
 QY 1361 CTACATCTCTGAGCCTTAACATACATGCTGTGGGTATCACACAGTGTGAGTGTCC 1420
 Db 61 CTACATCTCTGAGCCTTAACATACATGCTGTGGGTATCACACAGTGTGAGTGTCC 120
 QY 1421 AACCGTGTCAACAG 1480
 Db 121 AACCGTGTCAACAG 180
 QY 1481 AACCTATGTGAGCAGACATCCACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1540
 Db 181 AACCTATGTGAGCAGACATCCACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
 QY 1541 GATCTTCTGCTGTGGCAGACTTCAGAGACTTGTGGCTGAGGCGCTATTAGACAGAC 1600
 Db 241 GATCTTCTGCTGTGGCAGACTTCAGAGACTTGTGGCTGAGGCGCTATTAGACAGAC 300
 QY 1601 ACAGTATCAGTGAATGATCCATAAAGCTCCCTGTCACATCTTGGCCCAATGGGAATG 1660
 Db 301 ACAGTATCAGTGAATGATCCATAAAGCTCCCTGTCACATCTTGGCCCAATGGGAATG 360
 QY 1661 GATCTTCAACAAGAGCTACACAGCATTTTCCACAGAGATGC 1703
 Db 361 GATCTTCAACAAGAGCTACACAGCATTTTCCACAGAGATGC 403
 XX
 RESULT 8
 ABR54724
 ID ABR54724 standard; cDNA; 517 BP.
 XX
 AC ABR54724;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Human colon cancer-associated cDNA, SEQ ID No 194.
 XX
 KM Human; colon cancer; immunogenic; vaccine; tumour; gene; SS.
 XX
 OS Homo sapiens.
 XX
 PN WO200212280-A2.
 XX
 PD 14-FEB-2002.

XX
 PF 30-JUL-2001; 2001WO-US23826.
 XX
 PR 03-AUG-2000; 2000US-223265P.
 PR 02-OCT-2000; 2000US-237406P.
 PR 20-MAR-2001; 2001US-277495P.
 PR 03-JUL-2001; 2001US-302702P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Pyle RA, Xu J, Secrist H;
 XX
 DR WPI; 2002-257462/30.
 XX
 PT Novel polynucleotide encoding colon tumour polypeptides, useful as
 XX vaccines for treating colon cancers
 PS Claim 1; Page 206; 425pp; English.
 CC
 CC The invention relates to isolated polynucleotides (I) encoding colon
 CC tumour polypeptides (II). (I) is useful for stimulating an immune
 CC response in a patient and treating colon cancer in a patient.
 CC Oligonucleotides derived from (I) are useful for determining the presence
 CC of cancer in a patient. (I) and (II) are useful in pharmaceutical
 CC compositions, e.g. vaccines, and other compositions for the diagnosis
 CC and treatment of colon cancer. A composition comprising a first component
 CC selected from physiologically acceptable carriers and immunostimulants,
 CC and an antigen-presenting cell expressing (II) is useful for inhibiting
 CC development of cancer in a patient. (I) is useful in the design and
 CC preparation of ribozyme molecules for inhibiting expression of tumour
 CC polypeptides and (I). ABR54531-ABR54564 represent human colon cancer cDNA
 CC sequences of the invention.
 CC
 SO Sequence 517 BP; 121 A; 130 C; 124 G; 142 T; 0 other;
 XX
 Query Match 19.7%; Score 403; DB 24; Length 517;
 Best Local Similarity 100.0%; Pred. No. 2.6e-189;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1301 GGCTTGGTCCACCTGCTGTCAGCTCAGTCACTTCTGTAATGCTTGTGAGCCTTGC 1360
 Db 13 GGGTTTGGTCCACCTGCTGTCAGCTCAGTCACTTCTGTAATGCTTGTGAGCCTTGC 72
 QY 1361 CTACATCTCTGAGCCTTAACATACATGCTGTGGGTATCACACAGTGTGAGTGTCC 1420
 Db 73 CTACATCTCTGAGCCTTAACATACATGCTGTGGGTATCACACAGTGTGAGTGTCC 132
 QY 1421 AACCGTGTCAACAG 1480
 Db 133 AACCGTGTCAACAG 192
 QY 1481 AACCTATGTGAGCAGACATCCACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1540
 Db 193 AACCTATGTGAGCAGACATCCACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 252
 QY 1541 GATCTTCTGCTGTGGCAGACTTCAGAGACTTGTGGCTGAGGCGCTATTAGACAGAC 1600
 Db 253 GATCTTCTGCTGTGGCAGACTTCAGAGACTTGTGGCTGAGGCGCTATTAGACAGAC 312
 QY 1601 ACAGTATCAGTGAATGATCCATAAAGCTCCCTGTCACATCTTGGCCCAATGGGAATG 1660
 Db 313 ACAGTATCAGTGAATGATCCATAAAGCTCCCTGTCACATCTTGGCCCAATGGGAATG 372
 QY 1661 GATCTTCAACAAGAGCTACACAGCATTTTCCACAGAGATGC 1703
 Db 373 GATCTTCAACAAGAGCTACACAGCATTTTCCACAGAGATGC 415
 XX
 RESULT 9
 AAC76156
 ID AAC76156 standard; cDNA; 2988 BP.
 XX
 AC AAC76156;

XX 08-FEB-2001 (first entry)
XX
XX
XX Human OREFX ORF1711 polynucleotide sequence SEQ ID NO:3421.
XX
XX
XX Human; open reading frame; OREFX; detection; cytostatic; hepatotropic;
KW vulnary; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thromolytic; coagulant; vasotrophic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antihypertensive;
KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
XX Homo sapiens.
XX
XX MO200058473-A2.
XX
XX PD 05-OCT-2000.
XX
XX PF 31-MAR-2000; 2000MO-US08621.
XX
XX PR 31-MAR-1999; 99US-0127607.
XX PR 02-APR-1999; 99US-0127636.
XX PR 05-APR-1999; 99US-0127728.
XX PR 30-MAR-2000; 2000US-0540763.
XX PA
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Leach M;
PI WPI: 2000-602362/57.
DR P-PSDB: AAB41947.
DR
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 5; Page 2597-2599; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human OREFX open reading frames 1 to 3161. The OREFX
CC sequences have activities such as: cytosstatic; hepatotropic; vulnary;
CC antiparatic; antiparkinsonian; neurotropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thromolytic; coagulant; vasotrophic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an OREFX-associated disorder. The
CC nucleic acids can be used to express OREFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 2988 BP; 613 A; 916 C; 842 G; 617 T; 0 other;
SQ

Query Match Best Local Similarity 15.0%; Score 306; DB 21; Length 2988;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	452	TTTATATGCTTCAAGAAACCGTGGTCCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAAC	511
Db	1480	TTTTATGCTCAACATGGAACCTGGTCCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAAC	1530
QY	512	AGCCGGGCCCCGTGTTTGTGCACCTGCCCTGTACATCATCCACAGATGAATCATGCC	571
Db	1540	AGCCGGGCCCCGTGTTTGTGCACCTGCCCTGTACATCATCCACAGATGAATCATGCC	1590
QY	572	CGGGCTCACTGACAGGCTCTGTGCAGTCAACAGCCCTTTGAGTGGTGGAGAGCCCTGC	631
Db	1600	CGGGCTCACTGACAGGCTCTGTGCAGTCAACAGCCCTTTGAGTGGTGGAGAGCCCTGC	1650
QY	632	CGGCTCTCAACAGCCAGTGGTGCCTCAAGAGGAGCGCTTCTTCAACCTGAGTCCCTC	691
Db	1660	CGGCTCTCAACAGCCAGTGGTGCCTCAAGAGGAGCGCTTCTTCAACCTGAGTCCCTC	1710
QY	692	CCGCTGCTGAAGAAGACCCCTCCCTCAACCTGCATATGTGCACCTGGTCCGGAGCCCGG	751
Db	1720	CCGCTGCTGAAGAAGACCCCTCCCTCAACCTGCATATGTGCACCTGGTCCGGAGCCCGG	1770
QY	752	GCCGCTG 757	
Db	1780	GCCGCTG 1785	
RESULT 10			
ID	ABK54794		
ID	ABK54794	standard; cDNA; 389 BP.	
AC	ABK54794;		
DT	18-JUN-2002	(first entry)	
DE	Human colon cancer-associated cDNA, SEQ ID NO 264.		
XX	Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.		
KW	Homo sapiens.		
OS	WO200212280-A2.		
XX	14-FEB-2002.		
PN	30-JUL-2001; 2001WO-US23826.		
PF	03-AUG-2000; 2000US-223265P.		
PR	02-OCT-2000; 2000US-237406P.		
PR	20-MAR-2001; 2001US-277495P.		
PR	03-JUL-2001; 2001US-302702P.		
XX	(CORI-) CORIXA CORP.		
PA	Pyle RA, Xu J, Secretist H;		
PI	WPI; 2002-257462/30.		
DR			
XX			
XX			
PT	Novel polynucleotide encoding colon tumour polypeptides, useful as		
PT	vaccines for treating colon cancers		
XX	Claim 1; Page 225; 425pp; English.		
XX			
CC	The invention relates to isolated polynucleotides (I) encoding colon		
CC	tumour polypeptides (II). (I) is useful for stimulating an immune		
CC	response in a patient and treating colon cancer in a patient.		
CC	Oligonucleotides derived from (I) are useful for determining the presence		
CC	of cancer in a patient. (I) and (II) are useful in pharmaceutical		
CC	compositions, e.g. vaccines, and other compositions for the diagnosis		
CC	and treatment of colon cancer. A composition comprising a first component		
CC	selected from physiologically acceptable carriers and immunostimulants,		
CC	and an antigen-presenting cell expressing (II) is useful for inhibiting		
CC	development of cancer in a patient. (I) is useful in the design and		
CC	preparation of fibrozyme molecules for inhibiting expression of tumour		

CC polypeptides and (1). ABK54531-ABK55464 represent human colon cancer cDNA
 CC sequences of the invention.
 XX
 SQ Sequence 389 BP; 121 A; 83 C; 76 G; 109 T; 0 other;
 Query Match 5.6%; Score 114; DB 24; Length 389;
 Best Local Similarity 100.0%; Pred. No. 6.9e-46;
 Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1930 TGGCAGCTGTTAATAAATCCCAATAGGTTCTGTTAGAAATGCCCTTTTATG 1989
 DB 266 TGGCAGCTGTTAATAAATCCCAATAGGTTCTGTTAGAAATGCCCTTTTATG 325
 QY 1990 CTTCCTTAATATTAGACAGTAATGTCATTGTTATGGATCCCTAAAAA 2043
 DB 326 CTTCCTTAATATTAGACAGTAATGTCATTGTTATGGATCCCTAAAAA 379
 RESULT 11
 AAK92588/c
 ID AAK92588 standard; cDNA; 548 BP.
 AC AAK92588;
 XX
 DT 06-NOV-2001 (first entry)
 DE Human cDNA 3'-end sequence, SEQ ID NO: 1048.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 OS Homo sapiens.
 XX
 PN EP1130094-A2.
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-0114089.
 XX
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 PA (HELI-) HELIX RES INSTR.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 PI WPI; 2001-524255/58.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 XX
 PS Claim 3; SEQ ID NO 1048; 1380pp + sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is the nucleotide
 CC sequence of the 3'-end of a cDNA provided in the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 CC
 SQ Sequence 548 BP; 160 A; 109 C; 129 G; 148 T; 2 other;
 Query Match 5.1%; Score 105; DB 22; Length 548;
 Best Local Similarity 100.0%; Pred. No. 2e-41;
 Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1930 TGGCAGCTGTTAATAAATCCCAATAGGTTCTGTTAGAAATGCCCTTTTATG 1989
 DB 108 TGGCAGCTGTTAATAAATCCCAATAGGTTCTGTTAGAAATGCCCTTTTATG 49
 QY 1990 CTTCCTTAATATTAGACAGTAATGTCATTGTTATGGATCCCTAAA 2034
 DB 48 CTTCCTTAATATTAGACAGTAATGTCATTGTTATGGATCCCTAAA 4
 RESULT 12
 AAD24670
 ID AAD24670 standard; cDNA; 1647 BP.
 AC AAD24670;
 XX
 DT 12-MAR-2002 (first entry)
 DE Human drug metabolizing enzyme (DME)-5 cDNA.
 XX
 KW Human; drug metabolizing enzyme; gene therapy; autoimmune disorder;
 KW inflammatory disorder; acquired immune deficiency syndrome; infection;
 KW AIDS; adult respiratory distress syndrome; cell proliferative disorder;
 KW allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;
 KW cancer; endocrine disorder; hypothyroidism; pituitary disorder;
 KW gastrointestinal disorder; metabolic disorder; developmental disorder;
 KW liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;
 KW golfer; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
 KW DME-5; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key location/Qualifiers
 FT 1..1188
 FT CDS
 FT /tag- a
 FT /product- "Human DME-5 protein"
 FT 1..96
 FT sig-peptide
 FT /tag- b
 FT 97..1185
 FT mat-peptide
 FT /tag- c
 FT /product- "Human mature DME-5 protein #1"
 FT 1..105
 FT sig-peptide
 FT /tag- d
 FT 106..1185
 FT mat-peptide
 FT /tag- e
 FT /product- "Human mature DME-5 protein #2"
 FT MO200179468-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 12-APR-2001; 2001WO-US11869.
 XX
 PR 13-APR-2000; 2000US-197590P.
 PR 19-APR-2000; 2000US-198403P.
 PR 28-APR-2000; 2000US-200185P.
 PR 05-MAY-2000; 2000US-202234P.
 PR 11-MAY-2000; 2000US-203509P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Policky JL, Hafalla A, Burford N, Ring HZ, Lal P, Tribouley CM;
 PI Yao MG, Yue H, Tang YT, Patterson C, Das D, Sanjanwala MS;
 PI Gandhi AR, Reddy R, Khan FA, Baughn MR, Ramkumar J, Griffin JA;
 PI Au-Young J;
 PI WPI; 2002-066363/09.
 DR P-PDB; AAE15438.
 XX
 PT Novel isolated human drug metabolizing enzymes referred as DME 1-10,
 PT useful for diagnosing, treating, or preventing disorders associated
 PT with aberrant expression of DME such as allergy, anemia, asthma,
 PT infertility -
 XX

• Fri Aug 15 08:09:27 2003

us-09-645-078-1.ing

Page 13

Db 834 AGGCTGCGGCTCTACAGCCACGTGGTGCCTCAGAGAGTGGCTTCTTCACCTGCAG 892

Search completed: August 15, 2003, 04:50:47
Job time : 545 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2003, 04:13:39 ; Search time 3887 Seconds
(Without alignments)
12774.388 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043
Sequence: 1 gaattccattgttgttgta.....tgggacccaataaaaaaa 2043

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 1215238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 38370230

Minimum DB seq length: 300

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estdb:*
2: em_esthum:*
3: em_estlin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_hiv:*
19: em_gss_pln:*
20: em_gss_vrl:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	685	33.5	954	12	BI823850 603039012
2	579	28.3	583	9	BK490456 Homo sapi
3	534	26.1	668	9	AL709927 DKF2P8860
4	479	23.4	571	12	BM129080 1f17c04.y

Result No.	Score	Query Match	Length	ID	Description
5	322	15.8	593	12	BM969292 UI-CF-ENO
6	290	14.2	680	10	BE878439 MRO-ET010
7	288	14.1	669	13	BU684395 UT-CF-ENO
8	218	10.7	553	12	BM128370 1f13a12.x
9	218	10.7	553	12	BM128370 1f13a12.x
10	107	5.2	362	9	AM002418 w61903.x
11	105	5.1	417	9	AM572510 xq18g11.x
12	101	4.9	358	10	BF056840 7k10d12.x
13	74	3.6	419	9	AM572390 xuo9a06.x
14	59	2.9	525	10	BF197521 7b8a08.x
15	59	2.9	620	9	AI824100 w16c01.x
16	59	2.9	695	10	BE857538 7901a08.x
17	59	2.9	735	10	BE858652 7901a09.x
18	55	2.7	721	28	AQ741911 HS-5569.B
19	47	2.3	436	9	AM081348 xc41b06.x
20	27	1.3	571	12	BM246681 K0741E04-
21	27	1.3	571	12	BM245312 K0722H07-
22	27	1.3	695	12	BG964671 602831875
23	27	1.3	783	12	BG963298 602827716
24	27	1.3	852	12	BG966340 602832826
25	27	1.3	1923	11	AK009113 Mus muscu
26	23	1.1	362	10	BE340900 EST344988
27	23	1.1	397	9	AI345721 t8d4d04.x
28	23	1.1	687	14	CB167190 GWM602702
29	23	1.1	692	14	BY711808 BY711808
30	23	1.1	834	13	BO884877 AGCNCOURT
31	22	1.1	361	9	AI311365 t647h05.x
32	22	1.1	493	10	BF869376 t13-ET011
33	22	1.1	501	28	BH287165 CH230-113
34	22	1.1	516	9	AA689056 vs06c06.x
35	22	1.1	527	4	BX519156 RZPD Mus
36	22	1.1	910	13	BE603746 GS001 GS
37	22	1.1	1881	10	BE603746 GS001 GS
38	21	1.0	340	29	AG259978 LOCus Jap
39	21	1.0	410	28	AZ725813 RPT-24-9
40	21	1.0	424	29	AG234526 LOCus Jap
41	21	1.0	510	14	CB782649 AMGNMUC:S
42	21	1.0	541	28	AQ340290 HS-5026.A
43	21	1.0	616	28	AZ393547 IM0156H12
44	21	1.0	631	28	BH079110 RPT-24-2
45	21	1.0	633	28	AZ995163 2M0280G19

ALIGNMENTS

RESULT 1
BI823850
LOCUS
DEFINITION
603039012F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5179826 5',
mRNA sequence.
BI823850
VERSION
BI823850.1 GI:15935400
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
NIH-MGC http://mgc.nci.nih.gov/
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cga@fremail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM11448 row: h column: 03
High quality sequence stop: 856.

FEATURES
source

Location/Qualifiers

1. 954

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5179826"

/lab_host="DH10B"

/clone_1lb="NH_MGC_115"

/note="Organ: pooled brain, lung, testis; Vector: PCMV-SPORT6; Site: 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH_MGC Library."

BASE COUNT
ORIGIN

199 a 297 c 250 g 208 t

Query Match

33.5%; Score 685; DB 12; Length 954;

Best Local Similarity 100.0%; Pred. No. 1.1e-271;

Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 GGGCATGGCCGGCTAGCATGAGCTCTCAAAAGCAGGAGGAGCCCAAGCCACAAG 110
 DB 1 GGGCATGGCCGGCTAGCATGAGCTCTCAAAAGCAGGAGGAGCCCAAGCCACAAG 60
 QY 111 TCTTCCACTTCACACATGCTACTGCTTAAATAAAGCTCTGCTTTCTGTTT 170
 DB 61 TCTTCCACTTCACACATGCTACTGCTTAAATAAAGCTCTGCTTTCTGTTT 120
 QY 171 CCCAGATGGCATCTTGGCTCTATCTTCCATGTACAGCCACAATCAGCTCCCTGT 230
 DB 121 CCCAGATGGCATCTTGGCTCTATCTTCCATGTACAGCCACAATCAGCTCCCTGT 180
 QY 231 CTAATGAGGACAGCCGAGCGCATGAGTGTCTGTCTTCTGCTGCTGCTGCT 290
 DB 181 CTAATGAGGACAGCCGAGCGCATGAGTGTCTGTCTTCTGCTGCTGCTGCT 240
 QY 291 CTTCTTTTGGGGGAGCTTTTGGGAGCAGCCACCAATGTTTCTCATGATGAGACCCG 350
 DB 241 CTTCTTTTGGGGGAGCTTTTGGGAGCAGCCACCAATGTTTCTCATGATGAGACCCG 300
 QY 351 CCTGGCAGCTGTGATGATCCTTTCAGCAGACAGCCGCTGATGTGACATGGCTGTGC 410
 DB 301 CCTGGCAGCTGTGATGATCCTTTCAGCAGACAGCCGCTGATGTGACATGGCTGTGC 360
 QY 411 GGGATGTGATGAGGCGGCTTCTTGTGGACATGAGCGCTTGTGATGGCTACATGGAAC 470
 DB 361 GGGATGTGATGAGGCGGCTTCTTGTGGACATGAGCGCTTGTGATGGCTACATGGAAC 420
 QY 471 CTGTGTCGCCGAGACAGTCCAGGCTCTTTCAGTGGAGAAACAGCCGCGCTGTGTCG 530
 DB 421 CTGTGTCGCCGAGACAGTCCAGGCTCTTTCAGTGGAGAAACAGCCGCGCTGTGTCG 480
 QY 531 CACTGCTCTGTGATCATATCCACAAGATGAATATATCCCGCGGCTCACTGAGGCTCC 590
 DB 481 CACTGCTCTGTGATCATATCCACAAGATGAATATATCCCGCGGCTCACTGAGGCTCC 540
 QY 591 TGGGATGATGAGGCGCTTGTGAGGTGGAGAGGCGCTGCTCTACAGCAGCTGG 650
 DB 541 TGGGATGATGAGGCGCTTGTGAGGTGGAGAGGCGCTGCTCTACAGCAGCTGG 600
 QY 651 TGTCTAAGAGAGTGTGCTTTCACCTCAGTCCCTTACCGGCTGTGAAGACCCCT 710
 DB 601 TGTCTAAGAGAGTGTGCTTTCACCTCAGTCCCTTACCGGCTGTGAAGACCCCT 660
 QY 711 CCTCTAAGCTGATGTGACACT 735
 DB 661 CCTCTAAGCTGATGTGACACT 685

RESULT 2

HSM082755 standard; RNA; EST; 583 BP.

BX490456;

BX490456.1

09-MAY-2003 (Rel. 75, Created)

09-MAY-2003 (Rel. 75, Last updated, Version 1)

Homo sapiens mRNA; EST DKFZp686G0680_r1 (from clone DKFZp686G0680)

EST; expressed sequence tag.

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

[1]

Ansoerge W., Krieger S., Regiert T., Rittmuller C., Schwager B.,

Mewes H.W., Well B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;

Submitted (07-MAY-2003) to the EMBL/GenBank/DBJ databases.

M.P.S., Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY

This is the 5' sequence of the clone insert

CC Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de;

CC sequenced by EMBL (European Molecular Biology Laboratories),

CC Heidelberg/Germany) within the cDNA sequencing consortium of

CC the German Genome Project.

CC No s1 sequence available.

CC This clone (DKFZp686G0680) is available at the RZPD in Berlin.

CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

Key Location/Qualifiers

1. 583

/db_xref="taxon:9606"

/mol_type="mRNA"

/organism="Homo sapiens"

/clone="DKFZp686G0680"

/clone_1lb="686 (synonym: hlc33). Vector pSPORT1_sfi; host DH10B; sites SfiI + SfiIb"

/dev_stage="adult"

/tissue_type="cDNA-collection"

Sequence 583 BP; 120 A; 172 C; 154 G; 137 T; 0 other;
 Query Match 28.3%; Score 579; DB 2; Length 583;
 Best Local Similarity 100.0%; Pred. No. 6.8e-228;
 Matches 579; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 GAGGAGAAAAAGCCATGAGCCGCGGTAGCAGTGAAGCTCTCAAAAGCAGGAGGAGACCA 100
 DB 1 GAGGAGAAAAAGCCATGAGCCGCGGTAGCAGTGAAGCTCTCAAAAGCAGGAGGAGACCA 60
 QY 101 AGCCACAAGCTCTTCCACTTCAGCAGCAATGCTACTGCTTAAAAAATGAAGCTCCGTG 160
 DB 61 AGCCACAAGCTCTTCCACTTCAGCAGCAATGCTACTGCTTAAAAAATGAAGCTCCGTG 120
 QY 161 TTTTCTGTTTCCAGATGGCATCTTGGCTCTATTCTTCCACATGTACAGCCACAATC 220
 DB 121 TTTTCTGTTTCCAGATGGCATCTTGGCTCTATTCTTCCACATGTACAGCCACAATC 180
 QY 221 AGCTCCCTGTCTATGAGGACAGCCGAGGCGCATGAGTGTGCTGTCTCTCTG 280
 DB 181 AGCTCCCTGTCTATGAGGACAGCCGAGGCGCATGAGTGTGCTGTCTCTCTG 240

QY 281 CGCTGTGCTCTTCTTTTGTGGGACGCTTTTGGGAGCACCAGATGTTTCTACTG 340
 |||||
 Db 241 CGCTGTGCTCTTCTTTTGTGGGACGCTTTTGGGAGCACCAGATGTTTCTACTG 300
 |||||
 QY 341 ATGAGACCCGCTGGACGCTGTGGATGACCTTCAAGACAGACACGCTGTGATGCTGAC 400
 |||||
 Db 301 ATGAGACCCGCTGGACGCTGTGGATGACCTTCAAGACAGACACGCTGTGATGCTGAC 360
 |||||
 QY 401 ATGAGCTGTGGGATGATACGAGGCGCTCTTCTTGTGGACATGAGCGCTTGTGATGCC 460
 |||||
 Db 361 ATGAGCTGTGGGATGATACGAGGCGCTCTTCTTGTGGACATGAGCGCTTGTGATGCC 420
 |||||
 QY 461 TACATGACCTGTGCTCCGAGACAGTCCAGCCTTTTCAAGTGGAGAACAGCGGAGCC 520
 |||||
 Db 421 TACATGACCTGTGCTCCGAGACAGTCCAGCCTTTTCAAGTGGAGAACAGCGGAGCC 480
 |||||
 QY 521 CTGTGTCTGACCTGCTGTCATCATCCACAGATGAATCATCCCGGAGCTCAC 580
 |||||
 Db 481 CTGTGTCTGACCTGCTGTCATCATCCACAGATGAATCATCCCGGAGCTCAC 540
 |||||
 QY 581 TGCAGGCTCTGTGACATCAACAGCCCTTGTGAGGTGCTG 619
 |||||
 Db 541 TGCAGGCTCTGTGACATCAACAGCCCTTGTGAGGTGCTG 579
 |||||

RESULT 3
 AL709927 668 bp mRNA linear EST 12-JUN-2003
 LOCUS DKEP868602364_r1 666 (synonym: hlc3) Homo sapiens cDNA clone
 DEFINITION DKEP868602364 5', mRNA sequence.
 ACCESSION AL709927
 VERSION AL709927.1 GI:19693282
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 668)
 Bahr, A., Lauber, J., Mewes, H.W., Well, B., Amlid, C., Osanger, A., Fobo
 G., Han, M. and Wiemann, S.
 EST (Bahr, A., Lauber, J., Mewes, H.W., Well, B., et al.)
 TITLE Unpublished
 JOURNAL Contact: Bahr A
 COMMENT MIPs

Ingolstaedter Landstr. 1, D-85764 Neuherberg, Germany
 This is the 5' sequence of the clone insert.
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by Olgien (Hilden/Germany) within the cDNA sequencing
 consortium of the German Genome Project.
 No sl sequence available.
 This clone (DKEP868602364) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 FEATURES
 source
 1..668
 Location/Qualifiers

BASE COUNT 132 a 202 c 178 g 154 t 2 others
 ORIGIN
 Query Match 26.1%; Score 534; DB 9; Length 668;
 Best Local Similarity 100.0%; Pred. No. 2.3e-209;
 Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 AGAGAGAAAAAGCCATGAGCCCGGCTAGCACTGAGCCTCTCAAAAGCAGAGGAGCC 99
 |||||
 Db 6 AGAGAGAAAAAGCCATGAGCCCGGCTAGCACTGAGCCTCTCAAAAGCAGAGGAGCC 65
 |||||
 QY 100 AAGCCAAAGCTCTTCCACTTACACATGCTACTGCTTAAAAAATGAAGCTCTGCT 159
 |||||
 Db 66 AAGCCAAAGCTCTTCCACTTACACATGCTACTGCTTAAAAAATGAAGCTCTGCT 125
 |||||
 QY 160 GTTCTGTGTTCCAGATGAGCAGCTTGTGCTTATTTCTTCACATGTACAGCCACAT 219
 |||||
 Db 126 GTTCTGTGTTCCAGATGAGCAGCTTGTGCTTATTTCTTCACATGTACAGCCACAT 185
 |||||
 QY 220 CAGTCCCTGTCTTGAAGGACAGCCCGAGCATGAGCTGCTGTTCTTCTGCTG 279
 |||||
 Db 186 CAGTCCCTGTCTTGAAGGACAGCCCGAGCATGAGCTGCTGTTCTTCTGCTG 245
 |||||
 QY 280 GCGCTGTGCTCTTCTTGTGGGAGCTTTTGGGAGCAGCCAGATGTTTCTACT 339
 |||||
 Db 246 GCGCTGTGCTCTTCTTGTGGGAGCTTTTGGGAGCAGCCAGATGTTTCTACT 305
 |||||
 QY 340 GATGAGCCCGCTGCGACGCTGTGATGACCTTCAAGCAGACAGCCGCTGATGCA 399
 |||||
 Db 306 GATGAGCCCGCTGCGACGCTGTGATGACCTTCAAGCAGACAGCCGCTGATGCA 365
 |||||
 QY 400 CATGCTGTGCGGATGATGATGAGGCGGCTTCTTGTGGAGATGAGCGCTTGTATGC 459
 |||||
 Db 366 CATGCTGTGCGGATGATGATGAGGCGGCTTCTTGTGGAGATGAGCGCTTGTATGC 425
 |||||
 QY 460 CTACATGGAACCTGTGCTCCGAGACAGTCCAGCCTTTTCAAGTGGAGAACAGCGGCG 519
 |||||
 Db 426 CTACATGGAACCTGTGCTCCGAGACAGTCCAGCCTTTTCAAGTGGAGAACAGCGGCG 485
 |||||
 QY 520 CCTGTGTCTGACCTGCTGCTGTGACATCATCCCAAGATGAATCATCCCGG 573
 |||||
 Db 486 CCTGTGTCTGACCTGCTGCTGTGACATCATCCCAAGATGAATCATCCCGG 539
 |||||

RESULT 4
 BM129080 571 bp mRNA linear EST 12-MAR-2002
 LOCUS I17C04.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
 DEFINITION cDNA clone IMAGE:5676655 5' similar to TR:09YR3 09YR3
 N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE. ;, mRNA sequence.
 ACCESSION BM129080
 VERSION BM129080.1 GI:17123632
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 571)
 Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
 Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blustein, A.,
 Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
 M., Gibbons, M., McCann, R., Cole, R., Tsagarelashvili, R., Williams, T.,
 Jackson, Y. and Bowers, Y.
 Endocrine Pancreas Consortium
 Unpublished
 Other ESTs: I17C04.x1
 CONTACT: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@diolmp.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Juliana Brown
 (brown@fas.harvard.edu) This sequence now available from the IMAGE
 consortium, for clone orders contact: info@image.llnl.gov

FEATURES High quality sequence stop: 434.

Location/Qualifiers

1. 571

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5676655"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="Mellon Normalized Human Islet 4 M-HIS 1"

/note="Organ: Pancreas; Vector: pSPORT1; Site.1: Not 1;
Site.2: Sal 1; Starting library constructed using
SuperScript Plasmid Library Kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an Ecot of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxypatite chromatography and used to make this
library."

BASE COUNT 109 a 169 c 143 g 150 t
ORIGIN

Query Match 23.4%; Score 479; DB 12; Length 571;
Best Local Similarity 100.0%; Pred. No. 1.2e-186;

Matches 479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

106 CAGGCTTCACCTTCAGCACAATGCTACTGCTTAAATAAGTACCTCTGCTGTTCT 165
111
111
73 CAGGCTTCACCTTCAGCACAATGCTACTGCTTAAATAAGTACCTCTGCTGTTCT 132
111
111
166 GGTTCGCCAGATGGCCATCTGGCTCTATCTTCCACATGTACAGCCACAATCAGCTC 225
111
111
133 GGTTCGCCAGATGGCCATCTGGCTCTATCTTCCACATGTACAGCCACAATCAGCTC 192
111
111
226 CCGTCTATGAAGCAGACCCGAGGAGCAGTGTGCTCTCTTCCGCGGC 285
111
111
193 CCGTCTATGAAGCAGACCCGAGGAGCAGTGTGCTCTCTTCCGCGGC 252
111
111
286 TGGCTCTCTTTTGTGGGAGCTTTTGGGAGCAGCCAGATGTTTCTACCTGATGGA 345
111
111
253 TGGCTCTCTTTTGTGGGAGCTTTTGGGAGCAGCCAGATGTTTCTACCTGATGGA 312
111
111
346 GCGCGCGTGGACGCTGTGATGACCTTCAAGCAGACCCGCTGATGCTGACATGGC 405
111
111
313 GCGCGCGTGGACGCTGTGATGACCTTCAAGCAGACCCGCTGATGCTGACATGGC 372
111
111
406 TGTGGGAGATGATACGGGCGCTCTTGTGCGACATGAGGCTTTGATGCTACAT 465
111
111
373 TGTGGGAGATGATACGGGCGCTCTTGTGCGACATGAGGCTTTGATGCTACAT 432
111
111
466 GGAACCTGTGTCCCGGAGACAGTCCAGCTTTTCAAGTGGGAGAACAGCCGGCCCTGTG 525
111
111
433 GGAACCTGTGTCCCGGAGACAGTCCAGCTTTTCAAGTGGGAGAACAGCCGGCCCTGTG 492
111
111
526 TTCTGCACTGCTGTGACATCATCCACAGATGAATCATCCCGGGGCTCACTGCA 584
111
111
493 TTCTGCACTGCTGTGACATCATCCACAGATGAATCATCCCGGGGCTCACTGCA 551
111
111

RESULT 5

BM969292/c

LOCUS BM969292 593 bp mRNA linear EST 20-FEB-2003
DEFINITION UI-CF-ENO-acp-1-21-0-UI.s1 UI-CF-ENO Homo sapiens cDNA clone

ACCESSION BM969292
VERSION BM969292.1 GI:19586879

KEYWORDS

EST.
Homo sapiens (human)
Homo sapiens

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 593)
Bonaldo,M.F., Lennon,G. and Soares,M.B.

AUTHORS

Normalization and subtraction: two approaches to facilitate gene
discovery

TITLE

Genome Res. 6 (9), 791-806 (1996)

JOURNAL

97044477

MEDLINE

8889548

COMMENT

Contact: McCray, PB

McCrays Lab

University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171

Email:

paul-mccray@uiowa.edu

Tissue Procurement:

Dr. M. J. Welsh, University of Iowa

CDNA Library preparation:

Dr. M. Bento Soares, University of Iowa

CDNA Sequencing by:

Dr. M. Bento Soares, University of Iowa

Clone Distribution:

Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).

Seq primer:

M13 FORWARD

POLYA=Yes.

FEATURES

source

Location/Qualifiers

1. 593

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-CF-ENO-acp-1-21-0-UI"

/tissue_type="Primary Lung Cystic Fibrosis Epithelial
Cells"

/dev_stage="Adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone_lib="UI-CF-ENO"

/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site.1: EcoR I; Site.2: Not I;
UI-CF-ENO is a cDNA library containing the following
tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells.
The library was constructed according to Bonaldo, Lennon
and Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is CTGCTCAGGT.

TAG LIB-UI-CF-ENO

TAG_TISSUE=Human Lung Epithelial Cell Lines untreated LPS
6hr to LPS 24h

TAG_SEQ=CTGCTCAGGT

BASE COUNT 157 a 127 c 137 g 172 t
ORIGIN

Query Match 15.8%; Score 322; DB 12; Length 593;
Best Local Similarity 100.0%; Pred. No. 6.6e-122;

Matches 322; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1382 ACATGCTGTGGGATACACAGAGGTGAGTGTGTCACAGCTGCTCAGAGAGA 1441
111
111
593 ACATGCTGTGGGATACACAGAGGTGAGTGTGTCACAGCTGCTCAGAGAGA 534
111
111
1442 CTTTGTGTCATGCTTGTGTAGAAAACAGACTGGGACCTTATGTGACACACAT 1501
111
111
533 CTTTGTGTCATGCTTGTGTAGAAAACAGACTGGGACCTTATGTGACACACAT 474
111
111
1502 CCCACCATGAAACAGAGGATATGCTCTCTTTCTTGTGATCTCTCTGCGCAGAC 1561
111
111

Db 473 CCCACGAGTGAACAGGAGTATGCTCTTCTTTCTTTGATCTCTCTGCTGGAGAC 414
 QY 1562 TTGAGACATTTTGCGCTGGAGCCCTATTAGACACACAGTATGAGTAATGATC 1621
 Db 413 TTGAGACATTTTGCGCTGGAGCCCTATTAGACACACAGTATGAGTAATGATC 354
 QY 1622 CATTAACCTCCCTGCTCCACATCTTGCCCAATGGGAAATGATCTTCCACCAAGAGTCA 1681
 Db 353 CATTAACCTCCCTGCTCCACATCTTGCCCAATGGGAAATGATCTTCCACCAAGAGTCA 294
 QY 1682 CCAGCATTTTCCACAGAGATGC 1703
 Db 293 CCAGCATTTTCCACAGAGATGC 272

RESULT 6
 BF878439 680 bp mRNA linear EST 17-JAN-2001
 LOCUS BF878439
 DEFINITION BF878439
 ACCESSION BF878439
 VERSION BF878439.1 GI:12268569
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 680)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 10737800
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR06t2-MR0-ET0109-191100-002-h06&t3=2000-11-19&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 17
 High quality sequence stop: 641.
 Location/Qualifiers
 1. 680
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="ET0109"
 /note="Organ: lung_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 168 a 164 c 163 g 184 t 1 others
 ORIGIN
 Query Match 14.2%; Score 290; DB 10; Length 680;
 Best Local Similarity 100.0%; Pred. No. 9.9e-109;

Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1414 TGTGTCCACACGCTGCTCAGAGAAAGACTTTGTGTCATGCTTGTGCTAGAAACAG 1473
 Db 106 TGTGTCCACACGCTGCTCAGAGAAAGACTTTGTGTCATGCTTGTGCTAGAAACAG 165
 QY 1474 ACTGGGAACTTGTGAGACACATCCACCAGGAAACAGGATATGCTTCTTC 1533
 Db 166 ACTGGGAACTTGTGAGACACATCCACCAGGAAACAGGATATGCTTCTTC 225
 QY 1534 TTTCTTGATCTTCCTGCTGCGAGACTTCAGACTTTGCGCTGGAGCCCTATTTAA 1593
 Db 226 TTTCTTGATCTTCCTGCTGCGAGACTTCAGACTTTGCGCTGGAGCCCTATTTAA 285
 QY 1594 GCACGACACAGTATCAGTGGAAATGATCCATTAACCTCCCTGTCACATCTTGCCCAATG 1653
 Db 286 GCACGACACAGTATCAGTGGAAATGATCCATTAACCTCCCTGTCACATCTTGCCCAATG 345
 QY 1654 GGAATGATCTTCCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
 Db 346 GGAATGATCTTCCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 395

RESULT 7
 BU684395/c 669 bp mRNA linear EST 07-OCT-2002
 LOCUS BU684395
 DEFINITION BU684395
 ACCESSION BU684395
 VERSION BU684395.1 GI:23537302
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 669)
 Bernaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 8889548
 COMMENT Contact: McCray, PB
 McCray Lab
 University of Iowa
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
 Tel: 319 356 4866
 Fax: 319 356 7171
 Email: paul-mccray@uiowa.edu
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA library arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com) or from Open Biosystems (www.openbiosystems.com).
 Seq primer: M13 FORWARD
 POLYA=yes
 Location/Qualifiers
 1. 669
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_lib="UI-CF-ENO-aco-f-08-0-UI-3"
 /tissue_type="Primary Lung Cystic Fibrosis Epithelial Cells"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-CF-ENO"
 /note="Organ: Lung; Vector: pRT3-Pac (Pharmacia) with a modified Polylinker; Site_1: EcoR I; Site_2: Not I; UI-CF-ENO is a cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells."

The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)₁₈ tail. The sequence tag for this library is CTGCTCAGGT.

TAG: TISSUE=Human Lung Epithelial Cell Lines untreated LPS
TAG: LIB=UI-CF-ENO
6hr to LPS 24h

BASE COUNT 191 a 138 c 147 g 193 t
TAG_SRO=CTGCTCAGGT

Query Match 14.1%; Score 288; DB 13; Length 669;
Best Local Similarity 100.0%; Pred. No. 6.7e-108;
Matches 288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1416 TGTCCACAGCTGCTCAGCAGAGAGACTTTGTGTCCATGCTGTGTAGAAACAGAC 1475
DB 636 TGTCCACAGCTGCTCAGCAGAGAGACTTTGTGTCCATGCTGTGTAGAAACAGAC 577
QY 1476 TGGGGAACCTTATGTAGACAGCAGATCCACAGTGAACAGGATTTGCTCTTCTT 1535
DB 576 TGGGGAACCTTATGTAGAGCAGCAGATCCACAGTGAACAGGATTTGCTCTTCTT 517
QY 1536 TTCTTGATCTTCCTGCTGCGGAGACTTCAGAGACTTTGGCTGAGAGCCATTAAAG 1595
DB 516 TTCTTGATCTTCCTGCTGCGGAGACTTCAGAGACTTTGGCTGAGAGCCATTAAAG 457
QY 1596 AGCAGACAGTATCAGTGAATTCATTAACCTCCGTCCACATCTTGCCCAATGGG 1655
DB 456 AGCAGACAGTATCAGTGAATTCATTAACCTCCGTCCACATCTTGCCCAATGGG 397
QY 1656 GAATGATCTTCACCAAGAGCTCCAGCATTTTCCACAGAGATGC 1703
DB 396 GAATGATCTTCACCAAGAGCTCCAGCATTTTCCACAGAGATGC 349

RESULT 8
BM128370/c 553 bp mRNA 1linear EST 12-MAR-2002
LOCUS 1f11d12.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
DEFINITION CDNA clone IMAGE:567635 3', mRNA sequence.
ACCESSION BM128370
VERSION BM128370.1 GI:17122922
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 553)
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, R., Lemishka, I., Pearce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistein, A., Schmitt, A., Theising, B., Riter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagariswill, R., Williams, T., Jackson, Y. and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biochem.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center for information on

obtaining a clone please contact: Juliana Brown (brownj@fas.harvard.edu) This sequence now available from the IMAGE Consortium, for clone orders contact: info@image.llnl.gov
High quality sequence stop: 425.
Location/Qualifiers

1.553
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:567635"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/note="Organ: Pancreas; Vector: pSPORT1; Site: 1; Not 1; site:2; Sal 1; Starting library constructed using Superscript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoR of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 163 a 111 c 127 g 152 t
ORIGIN

Query Match 10.7%; Score 218; DB 12; Length 553;
Best Local Similarity 100.0%; Pred. No. 5.3e-79;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1486 TATGTAGCAGCAGCATCCACAGGTGAACAGGATATGCTCTCTCTTTTGTATCT 1545
DB 553 TATGTAGCAGCAGCATCCACAGGTGAACAGGATATGCTCTCTCTTTTGTATCT 494
QY 1546 TTCTGCTGCGGAGACTTCAGAGACTTTGGCTGAGAGCCATTAAAGCAGACAGT 1605
DB 493 TTCTGCTGCGGAGACTTCAGAGACTTTGGCTGAGAGCCATTAAAGCAGACAGT 434
QY 1606 ATCAGTGAATTCATTAACCTCCGTCCACATCTTGCCCAATGGGAATGATCT 1665
DB 433 ATCAGTGAATTCATTAACCTCCGTCCACATCTTGCCCAATGGGAATGATCT 374
QY 1666 TTACCAAGAGCTCCAGCATTTTCCACAGAGATGC 1703
DB 373 TTACCAAGAGCTCCAGCATTTTCCACAGAGATGC 336

RESULT 9
BM128831/c 553 bp mRNA 1linear EST 12-MAR-2002
LOCUS 1f11c04.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
DEFINITION CDNA clone IMAGE:567655 3', mRNA sequence.
ACCESSION BM128831
VERSION BM128831.1 GI:17123383
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 553)
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, R., Lemishka, I., Pearce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistein, A., Schmitt, A., Theising, B., Riter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagariswill, R., Williams, T., Jackson, Y. and Bowers, Y.
Endocrine Pancreas Consortium

JOURNAL COMMENT

Unpublished
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557

Email: dmelton@hobp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center for information on
obtaining a clone please contact: Juliana Brown
(brownjia.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
High quality sequence stop: 425.

FEATURES

SOURCE

Location/Qualifiers
1. 553
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:567655"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/note="Organ: Pancreas; Vector: pSPOR1; Site: 1; Not 1;
Site 2: Sal 1; Starting library constructed using
SuperScript Plasmid Library kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an Ecot of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."

BASE COUNT 163 a 111 c 127 g 152 t
ORIGIN

Query Match 10.7%; Score 218; DB 12; Length 553;
Best Local Similarity 100.0%; Pred. No. 5.3e-79;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1486 TATGTGACGACATCCACAGAGGATTTGCTCTCTTTCTTGATCT 1545
|||||
553 TATGTGACGACATCCACAGAGGATTTGCTCTCTTTCTTGATCT 494
1546 TCCGTCTGGGAGCTCAGAGACTTGGCCCTGAGGCGCTATTAGCAGACAGT 1605
|||||
493 TCCGTCTGGGAGCTCAGAGACTTGGCCCTGAGGCGCTATTAGCAGACAGT 434
1606 ATCAGTGAATTCATTAACCTCCTGTCCACATTTGGCCATGGGATGATCT 1665
|||||
433 ATCAGTGAATTCATTAACCTCCTGTCCACATTTGGCCATGGGATGATCT 374
1666 TTCACCAAGAGCTCAGCAGATTTTCCACAGAGATGC 1703
|||||
373 TTCACCAAGAGCTCAGCAGATTTTCCACAGAGATGC 336

RESULT 10
AM002418/c 362 bp mRNA linear EST 27-OCT-1999
LOCUS WU1903.x1 NCI_CGAP_G6C Homo sapiens cDNA clone IMAGE:2524564 3',
DEFINITION mRNA sequence.
ACCESSION AM002418
VERSION AM002418.1 GI:5849334
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 362)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished

JOURNAL COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/BLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco.

FEATURES

SOURCE

Location/Qualifiers
1. 362
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2524564"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_G6C"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site: 1; Not 1; Site 2: Eco RI; Plasmid DNA
from the normalized library NCI-CGAP_G6C was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clones
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 108 a 69 c 77 g 108 t
ORIGIN

Query Match 5.2%; Score 107; DB 9; Length 362;
Best Local Similarity 100.0%; Pred. No. 3.7e-33;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1930 TCCCATCTCTTAATCTAAATCCCAATAGCTTCTGTAGAAATCCCTTTTANG 1989
|||||
107 TCCCATCTCTTAATCTAAATCCCAATAGCTTCTGTAGAAATCCCTTTTANG 48
1990 CTCTTAATTTATACATAATGTTCAATTTTATGAGATCCTATAAA 2036
|||||
47 CTCTTAATTTATACATAATGTTCAATTTTATGAGATCCTATAAA 1

RESULT 11
AM572510/c 417 bp mRNA linear EST 13-MAR-2000
LOCUS XG18911.x2 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2751044 3',
DEFINITION mRNA sequence.
ACCESSION AM572510
VERSION AM572510.1 GI:7237243
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 417)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
JOURNAL COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

```

/organism="Homo sapiens"
/mol_type="mRNA"

```

Average insert size 1.7 kb. Life Technologies catalog #: 15930-016.

BASE COUNT	122	78	89	130
	11531-019"			

ORIGIN

Query Match 3.6%; Score 74; DB 9; Length 419;
 Best Local Similarity 100.0%; Pred. No. 1.4e-19;
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1630 TCCCTGTCACACATCTTCCCAATGGGAGATGATCTTTCACCAAGAGCTCACCAGCAT 1689
 Db 419 TCCCTGTCACACATCTTCCCAATGGGAGATGATCTTTCACCAAGAGCTCACCAGCAT 360
 QY 1690 TTCACAGAGATGC 1703
 Db 359 TTCACAGAGATGC 346

RESULT 14
 BF197521/c
 LOCUS
 DEFINITION 525 bp mRNA linear EST 03-NOV-2000
 7684608.x1 NCI-CGAP_Kid11 Homo sapiens CDNA clone IMAGE:3642903 3'
 similar to TR:09UBY3 09UBY3 N-ACETYLGLUCOSAMINE
 6-O-SULFOTRANSFERASE. ; mRNA sequence.

ACCESSION BF197521.1 GI:11086670
 VERSION BF197521
 KEYWORDS EST.
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 525)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL, send email to:
 info@image.llnl.gov
 High quality sequence stop: 451.

FEATURES

source
 1. 525
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3642903"
 /lab_host="DH10B"
 /clone_idb="NCI-CGAP_Kid11"
 /note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 Plasmid DNA from the normalized library NCI-CGAP_Kid3 was
 prepared, and ss circles were made in vitro. Following HRP
 purification, this DNA was used as tracer in a subtractive
 hybridization reaction. The driver was PCR-amplified cDNAs
 from a pool of 5,000 clones made from the same library
 (clones 1322376-1323911, 1456007-1456775, and
 1500552-1502855). Subtraction by Bento Soares and M.
 Fatima Bonaldo."
 BASE COUNT 90 a 163 c 187 g 85 t
 ORIGIN

Query Match 2.9%; Score 59; DB 10; Length 525;
 Best Local Similarity 100.0%; Pred. No. 2e-13;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTCCGCTCTACAGCAGCTGGTGTCTCAAGAGAGTGGCTTCTCAACTGCAG 682
 Db 260 AGGCTCCGCTCTCTACAGCAGCTGGTGTCTCAAGAGAGTGGCTTCTCAACTGCAG 202

RESULT 15
 A1824100/c
 LOCUS 620 bp mRNA linear EST 21-DEC-1999
 w146c01.x1 NCI-CGAP_Lu19 Homo sapiens CDNA clone IMAGE:2405856 3'
 similar to TR:075667 075667 DJ7116.4 ; contains PTRS.b2 PTR5
 repetitive element ; mRNA sequence.

ACCESSION A1824100
 VERSION A1824100
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 620)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www-bio.llnl.gov/dbp/image/image.html
 Insert Length: 917 Std Error: 0:00
 Seq primer: -40UP from GAPCO
 High quality sequence stop: 490.

FEATURES

source
 1. 620
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2405856"
 /tissue_type="squamous cell carcinoma, poorly
 differentiated (4 pooled tumors, including primary and
 metastatic)."
 /dev_stage="adult"
 /lab_host="DH10B (phage-resistant)"
 /clone_idb="NCI-CGAP_Lu19"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 pooled lung tumor tissue, and was then primed with a Not I
 - oligo(dT) primer. Double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. Library went through one round of
 normalization. Library constructed by Bento Soares and M.
 Fatima Bonaldo."
 BASE COUNT 106 a 193 c 227 g 89 t 3 others
 ORIGIN

Query Match 2.9%; Score 59; DB 9; Length 620;
 Best Local Similarity 100.0%; Pred. No. 1.9e-13;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTCCGCTCTACAGCAGCTGGTGTCTCAAGAGAGTGGCTTCTCAACTGCAG 682
 Db 314 AGGCTCCGCTCTCTACAGCAGCTGGTGTCTCAAGAGAGTGGCTTCTCAACTGCAG 256

Search completed: August 15, 2003, 07:57:24
 Job time : 3892 secs

THIS PAGE BLANK (insert)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2003, 04:41:40 ; Search time 142 Seconds
(without alignments)
6350.322 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043

Sequence: 1 gaattccatgtgtgtgtgta.....tgggacccaataaaaaaa 2043

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 241664

Minimum DB seq length: 300

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/lna/PCTUS.COMB.seq: *
6: /cgn2_6/ptodata/1/lna/Backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1547	75.7	2032	3	US-09-045-284A-1
2	1547	75.7	2032	4	US-09-180-911-2
3	1153	56.4	2065	2	US-09-786-240-26
4	19	0.9	3792	2	US-08-992-334-1
5	19	0.9	3792	2	US-08-302-752-1
6	19	0.9	5234	2	US-08-992-334-2
7	19	0.9	5234	2	US-08-302-752-2
8	19	0.9	6722	2	US-08-992-334-3
9	19	0.9	6722	3	US-08-302-752-3
10	18	0.9	536	3	US-09-020-956-69
11	18	0.9	536	3	US-09-030-607-69
12	18	0.9	536	4	US-09-439-313-69
13	18	0.9	536	4	US-09-352-616A-69
14	18	0.9	536	4	US-09-232-149A-69
15	18	0.9	639	4	US-09-107-532A-1333
16	18	0.9	822	3	US-09-020-956-29
17	18	0.9	822	3	US-09-030-607-29
18	18	0.9	822	4	US-09-439-313-29
19	18	0.9	822	4	US-09-352-616A-29
20	18	0.9	822	4	US-09-232-149A-29
21	18	0.9	1197	4	US-09-257-179-26
22	18	0.9	1368	4	US-09-118-554-61
23	18	0.9	1368	4	US-09-118-627-61
24	18	0.9	1368	4	US-09-602-877A-61
25	18	0.9	1913	4	US-09-599-360B-40
26	18	0.9	1927	4	US-09-336-536-66
27	18	0.9	4351	4	US-09-634-238-25

28	17	0.8	673	4	US-09-280-116-93	Sequence 93, App1
29	17	0.8	813	4	US-09-252-991A-1600	Sequence 1600, App
30	17	0.8	857	4	US-09-171-209-4	Sequence 4, App1
31	17	0.8	1338	4	US-09-328-352-1988	Sequence 1988, App
32	17	0.8	1466	3	US-09-130-242-8	Sequence 8, App1
33	17	0.8	1524	4	US-09-252-991A-1651	Sequence 1651, App
34	17	0.8	1540	1	US-08-286-872-5	Sequence 5, App1
35	17	0.8	1572	4	US-09-612-964-1	Sequence 1, App1
36	17	0.8	1575	1	US-07-988-260B-2	Sequence 2, App1
37	17	0.8	1683	4	US-09-252-991A-11226	Sequence 11226, App
38	17	0.8	1764	1	US-08-036-555B-150	Sequence 150, App
39	17	0.8	1764	1	US-08-469-569-150	Sequence 150, App
40	17	0.8	1764	1	US-08-249-322A-150	Sequence 150, App
41	17	0.8	1764	1	US-08-469-526A-150	Sequence 150, App
42	17	0.8	1764	2	US-08-734-591A-150	Sequence 150, App
43	17	0.8	1764	2	US-08-469-660-150	Sequence 150, App
44	17	0.8	1764	3	US-08-341-018-57	Sequence 57, App1
45	17	0.8	1764	3	US-08-470-335-150	Sequence 150, App

ALIGNMENTS

RESULT 1
US-09-045-284A-1
Sequence 1, Application US/09045284A
Patent No. 6265192

GENERAL INFORMATION:
APPLICANT: Bistrup, Annette
APPLICANT: Rosen, Steven D.
APPLICANT: Hemmerich, Stefan
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
FILE REFERENCE: 6510-107U51
CURRENT APPLICATION NUMBER: US/09/045, 284A
CURRENT FILING DATE: 1998-03-20
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2032
TYPE: DNA
ORGANISM: Homo sapiens
US-09-045-284A-1

Query Match 75.7%; Score 1547; DB 3; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	106	CAAGCTCTCCACTTACGACACATGCTACTGCTAAATAAGAGCTCTGCTGTTCT	165	Sequence 1, App1
DB	160	CAAGCTCTCCACTTACGACACATGCTACTGCTAAATAAGAGCTCTGCTGTTCT	219	Sequence 26, App1
QY	166	GGTTCCTCCAGTGGCCATCTTGGCTCTATCTTCACATGACGCCAATCAGCTC	225	Sequence 1, App1
DB	220	GGTTCCTCCAGTGGCCATCTTGGCTCTATCTTCACATGACGCCAATCAGCTC	279	Sequence 2, App1
QY	226	CTGTCTATGAAGGACAGCCGAGCATGACGTGCTGTTCTGTTCTGCGCTC	285	Sequence 3, App1
DB	280	CTGTCTATGAAGGACAGCCGAGCATGACGTGCTGTTCTGTTCTGCGCTC	339	Sequence 69, App1
QY	286	TGCGCTCTCTTGTGGGAGCTTTTGGGACACCCAGATGTTTCTTACTGATGA	345	Sequence 69, App1
DB	340	TGCGCTCTCTTGTGGGAGCTTTTGGGACACCCAGATGTTTCTTACTGATGA	399	Sequence 66, App1
QY	346	GCCCGCTGGCAGCTGTGATGACCTTCAAGCAGAGACGCCGCTGATGCGACATGGC	405	Sequence 29, App1
DB	400	GCCCGCTGGCAGCTGTGATGACCTTCAAGCAGAGACGCCGCTGATGCGACATGGC	459	Sequence 29, App1
QY	406	TGTGCGGATCTGATACGGGCGCTCTTGTGGAGATGACGCTTTGATGCTACAT	465	Sequence 26, App1
DB	460	TGTGCGGATCTGATACGGGCGCTCTTGTGGAGATGACGCTTTGATGCTACAT	519	Sequence 61, App1
QY	466	GGAACCTGTGTCCTCCGAGACATGTCACCTCTTCACTGGAGAAACGCCGCGCTGTG	525	Sequence 40, App1
				Sequence 66, App1
				Sequence 25, App1

Db 520 GGAACCTGGTCCCCGAGAGACAGTCAGCCTCTTCACTGGGAGAAAGCCGGGCGCTGTG 579
QY 526 TTCTGACCTGCGCTGTGACATCATCCACAAGATGAATCATCCCCGGGCTCACTGCAG 585
Db 580 TTCTGACCTGCGCTGTGACATCATCCACAAGATGAATCATCCCCGGGCTCACTGCAG 639
QY 586 GCTCCGTGTGAGTCAACAGCCCTTTAGAGTGTGTGAGAAAGGCGCTGCCGCTCTACAGCCA 645
Db 640 GCTCCGTGTGAGTCAACAGCCCTTTAGAGTGTGTGAGAAAGGCGCTGCCGCTCTACAGCCA 699
QY 646 CGTGGTGTCAAGAGAGTGGGCTTTCTCAACCTGCAGTCCCTTACCCGCTGTGAAAGA 705
Db 700 CGTGGTGTCAAGAGAGTGGGCTTTCTCAACCTGCAGTCCCTTACCCGCTGTGAAAGA 759
QY 706 CCCCTCCCTCAACCTGCATATCGTGCACCTGTGTCCGGAGCCCGGGGCGCTGTTCGTTTC 765
Db 760 CCCCTCCCTCAACCTGCATATCGTGCACCTGTGTCCGGAGCCCGGGGCGCTGTTCGTTTC 819
QY 766 CCGAGAGACGACAAAGAGAGATCTCATGATTCAGTCCGATTTGTATGGGGCAGCATGA 825
Db 820 CCGAGAGACGACAAAGAGAGATCTCATGATTCAGTCCGATTTGTATGGGGCAGCATGA 879
QY 826 GCAAAAACCTCAAGAGAGAGAGACCAACCTACTATGTGATGAGTGCATGTCCCAAGCCA 885
Db 880 GCAAGAACTCAAGAGAGAGAGACCAACCTACTATGTGATGAGTGCATGTCCCAAGCCA 939
QY 886 GCTGAGATCTCAAGAGACATCCAGTCTGTGCGCAAGGCGCTGCAGAAAGCTACCTGCT 945
Db 940 GCTGAGATCTCAAGAGACATCCAGTCTGTGCGCAAGGCGCTGCAGAAAGCTACCTGCT 999
QY 946 TGTGGCCCTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTTCCCGAATGATGAAAT 1005
Db 1000 TGTGGCCCTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTTCCCGAATGATGAAAT 1059
QY 1006 CGTGGATTTGGAATTTCTGCCCATCTTCAAGACCTGGGTGCATATCATCCAGAGCAA 1065
Db 1060 CGTGGATTTGGAATTTCTGCCCATCTTCAAGACCTGGGTGCATATCATCCAGAGCAA 1119
QY 1066 GGGCATAGGTGACACACGCTTTCACACAATGCGAGGATGCCCTTAATGTCTCCAGGC 1125
Db 1120 GGGCATAGGTGACACACGCTTTCACACAATGCGAGGATGCCCTTAATGTCTCCAGGC 1179
QY 1126 TTGGGCTGTGGTCTTGGCCATGAAAAAGTTCTGACATTCAGAAAGCGCTGGGCATGC 1185
Db 1180 TTGGGCTGTGGTCTTGGCCATGAAAAAGTTCTGACATTCAGAAAGCGCTGGGCATGC 1239
QY 1186 CATGAATTTGCTGGGTACCGCCACGCTGAGATCTGAACAAGAAAGAAAGCTGTGTCT 1245
Db 1240 CATGAATTTGCTGGGTACCGCCACGCTGAGATCTGAACAAGAAAGAAAGCTGTGTCT 1299
QY 1246 GGAATCTTGTCTACTGAGCTGTCCCTGAGCAATTCACATAAGAGGTTGAGAGGCTT 1305
Db 1300 GGAATCTTGTCTACTGAGCTGTCCCTGAGCAATTCACATAAGAGGTTGAGAGGCTT 1359
QY 1306 TGTGTCACCTGGTGTACAGCTCACTGTCATGTAATGCTTGAAGGCTTGGCTACAC 1365
Db 1360 TGTGTCACCTGGTGTACAGCTCACTGTCATGTAATGCTTGAAGGCTTGGCTACAC 1419
QY 1366 TGTGTCAGCCTTAACATACATGTCGTGGGTATCACACTGAGTGTGATGTCTACACAG 1425
Db 1420 TGTGTCAGCCTTAACATACATGTCGTGGGTATCACACTGAGTGTGATGTCTACACAG 1479
QY 1426 TGTGTCAGCAGAAAGACTTTTGTGTCCATGCTTGTCTTGAAGAAACAGACTGGGAACT 1485
Db 1480 TGTGTCAGCAGAAAGACTTTTGTGTCCATGCTTGTCTTGAAGAAACAGACTGGGAACT 1539
QY 1486 TATGTAGAGAGACATCCACAGTGAAGAGGTAATGCTCTTCTTTTATATCT 1545
Db 1540 TATGTAGAGAGACATCCACAGTGAAGAGGTAATGCTCTTCTTTTATATCT 1599
QY 1546 TCTGTCTGTGGAGAGACTTTCAGAGACTTTGTGGCTGAGAGCCATTAAAGCAGACAGT 1605
Db 1600 TCTGTCTGTGGAGAGACTTTCAGAGACTTTGTGGCTGAGAGCCATTAAAGCAGACAGT 1659

QY 1606 ATCAGTGAATTTATCATTAACCTCCCTGTCACATCTTGGCCAAATGGGAATGATCT 1665
Db 1660 ATCAGTGAATTTATCATTAACCTCCCTGTCACATCTTGGCCAAATGGGAATGATCT 1719
QY 1666 TTCACCAAGAGCTCACACAGATTTTCCACAGAGATGC 1703
Db 1720 TTCACCAAGAGCTCACACAGATTTTCCACAGAGATGC 1757

RESULT 2
US-09-190-911-2
; Sequence 2, Application US/09190911
; Patent No. 6365365
; GENERAL INFORMATION:
; APPLICANT: Blstrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/09/190,911
; EARLIER FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; EARLIER FILING DATE: 1998-03-20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens
US-09-190-911-2

Query Match 75.7%; Score 1547; DB 4; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 106 CAAGTCTTCCACTTACAGACAAATGCTACTGCTTAAAAAATGAAAGCTCTGCTGTTCT 165
Db 160 CAAGTCTTCCACTTACAGACAAATGCTACTGCTTAAAAAATGAAAGCTCTGCTGTTCT 219
QY 166 GGTTCACAGATGGGCACTTGTGCTATTTCTTCCAGATGTACAGGCACAAATCAGCTC 225
Db 220 GGTTCACAGATGGGCACTTGTGCTATTTCTTCCAGATGTACAGGCACAAATCAGCTC 279
QY 226 CCTGTATGAAGACACAGCCGAGGCGATGACAGTGTGCTGCTCTGCGGCTC 285
Db 280 CCTGTATGAAGACACAGCCGAGGCGATGACAGTGTGCTGCTCTGCGGCTC 339
QY 286 TGGCTCTTCTTTTGTGGGCACTTTTGTGGGACGACCCAGATGTTTCTACTGATGA 345
Db 340 TGGCTCTTCTTTTGTGGGCACTTTTGTGGGACGACCCAGATGTTTCTACTGATGA 399
QY 346 GCCCGCTGGACAGTGTGATGACCTTCAAGCAGAGACACGCGCTGTGATGCTGACATGGC 405
Db 400 GCCCGCTGGACAGTGTGATGACCTTCAAGCAGAGACACGCGCTGTGATGCTGACATGGC 459
QY 406 TGTGCGGAGATGATACGGGCGCTTCTTGTGCGACATGAGGCTTTTGATGCTACAT 465
Db 460 TGTGCGGAGATGATACGGGCGCTTCTTGTGCGACATGAGGCTTTTGATGCTACAT 519
QY 466 GGAACCTGTGTCGGGAGACAGTCCAGCTCTTTCAGTGGGAAACAGCCGGGCGCTGTG 525
Db 520 GGAACCTGTGTCGGGAGACAGTCCAGCTCTTTCAGTGGGAAACAGCCGGGCGCTGTG 579
QY 526 TTCTGACCTGCGCTGTGACATCATCCACAAGATGAATCATCCCCGGGCTCACTGCAG 585
Db 580 TTCTGACCTGCGCTGTGACATCATCCACAAGATGAATCATCCCCGGGCTCACTGCAG 639
QY 586 GCTCTGTGTGAGTCAACAGCCCTTTGAGTGTGTGAGAAAGGCGCTGCCGCTCTACAGCA 645
Db 640 GCTCTGTGTGAGTCAACAGCCCTTTGAGTGTGTGAGAAAGGCGCTGCCGCTCTACAGCA 699

OY	644	GTGTGTGCTCAAGGAGGGCGCTTCTTCAACCTGCAGTCCCTCTAACCCTGCTGTAAGA	705
Dp	700	CGTGTGTCTCAAGGAGGTGGCGCTTCTTCAACCTGCAGTCCCTCTAACCCTGCTGTAAGA	759
OY	706	CCCTCCCTCAACCTGCATATCTGACACTGTGCCGGAGCCCGGGCGCTGTCCGATC	765
Dp	760	CCCTCCCTCAACCTGCATATCTGACACTGTGCCGGAGCCCGGGCGCTGTCCGATC	819
OY	766	CCGAGAACGCACAAAGGAGATCTTCATGATGATGACAGTCCGATTGTGATGGGGCAGCATGA	825
Dp	820	CCGAGAACGCACAAAGGAGATCTTCATGATGATGACAGTCCGATTGTGATGGGGCAGCATGA	879
OY	826	GCAAAAACCTCAAGAGGAGGAGACCAACCTTACTATGTGATGGAGGTCATCTGCACAAAGCA	885
Dp	880	GCAGAAACCTCAAGAGGAGGAGACCAACCTTACTATGTGATGGAGGTCATCTGCACAAAGCA	939
OY	886	GCTGAGATCTACAAAGACCATCCAGTCTTGCCCAAGGCCCTGCAGAGCCTACCTGCT	945
Dp	940	GCTGAGATCTACAAAGACCATCCAGTCTTGCCCAAGGCCCTGCAGAGCCTACCTGCT	999
OY	946	TGTCGCGATATAGGACCTGGGCTCGAGGCCCTGTGGCCCGCAGACTTCCGAATGTATGATT	1009
Dp	1000	TGTCGCGATATAGGACCTGGGCTCGAGGCCCTGTGGCCCGCAGACTTCCGAATGTATGATT	1055
OY	1006	CGTGGGATTTGAAATCTTGCCCATCTTCAGACCTGGGTGCATAACATCACCCGAGGCA	1065
Dp	1060	CGTGGGATTTGAAATCTTGCCCATCTTCAGACCTGGGTGCATAACATCACCCGAGGCA	1119
OY	1066	GGGAGATGGGTACACCGCTTCCACACAAATGCCAGGATATCCCTTATGTCTCCAGGC	1122
Dp	1120	GGGAGATGGGTACACCGCTTCCACACAAATGCCAGGATATCCCTTATGTCTCCAGGC	1179
OY	1126	TTGGCGCGCTTCTTGCCCTATGAAAAGGTTCTCGATTCAGAAAGCCTGTGGCGATGC	1185
Dp	1180	TTGGCGCGCTTCTTGCCCTATGAAAAGGTTCTCGATTCAGAAAGCCTGTGGCGATGC	1233
OY	1186	CATGAATTTGCTGGGCTACCGCCACAGTCGATCTGACACAGAACAGAAACCTGTGCT	1245
Dp	1240	CATGAATTTGCTGGGCTACCGCCACAGTCGATCTGACACAGAACAGAAACCTGTGCT	1299
OY	1246	GGATCTTCTGCTACCTGGAGCTGTGCCGACAAATCCAGTAAGGGTTGGAAGGCTT	1305
Dp	1300	GGATCTTCTGCTACCTGGAGCTGTGCCGACAAATCCAGTAAGGGTTGGAAGGCTT	1355
OY	1306	TGCTGCCACCTGGTGTACGCTCAGTCACCTTCTCTGTAATGCTTCTGAGCCTTGCTACA	1365
Dp	1360	TGCTGCCACCTGGTGTACGCTCAGTCACCTTCTCTGTAATGCTTCTGAGCCTTGCTACA	1419
OY	1366	TCTGTGAGCCTTAATCAATGCTGTGGGTATTCACACTGATGTGATGTGTCTCACAGC	1422
Dp	1420	TCTGTGAGCCTTAATCAATGCTGTGGGTATTCACACTGATGTGATGTGTCTCACAGC	1479
OY	1426	TGCTCAAGCAGAAAGACTTTGTGTCTCACTGCTTGTGTAGAAAAACAGACTGGGGAACCT	1485
Dp	1480	TGCTCAAGCAGAAAGACTTTGTGTCTCACTGCTTGTGTAGAAAAACAGACTGGGGAACCT	1533
OY	1486	TATGTAGACAGACATCCCAACAGTGAACAGGGTATTGCTCTCTCTTCTTCTTGATCT	1545
Dp	1540	TATGTAGACAGACATCCCAACAGTGAACAGGGTATTGCTCTCTCTTCTTCTTGATCT	1599
OY	1546	TCCGTGTGGGCGAGACTTCAGAGACTTGTGGCCGTGGAGGCCCTTATAGCAGCAGACAGT	1609
Dp	1600	TCCGTGTGGGCGAGACTTCAGAGACTTGTGGCCGTGGAGGCCCTTATAGCAGCAGACAGT	1655
OY	1606	ATCAGTGAATTTGATTCATAAACCTCCCTCTCCACATCTTGCCCAATGGGGAATGATCT	1665
Dp	1660	ATCAGTGAATTTGATTCATAAACCTCCCTCTCCACATCTTGCCCAATGGGGAATGATCT	1719
OY	1666	TTTCACCAAAAGAGCTCACCGACATTTTTCACAGAGATGC 1703	
Dp	1720	TTTCACCAAAAGAGCTCACCGACATTTTTCACAGAGATGC 1757	

```

RESULT 3
US-09-786-240-26
: Sequence 26, Application US/09786240
: Patent No. 6558935
: GENERAL INFORMATION:
: APPLICANT: INCYTE PHARMACEUTICALS, INC.
: APPLICANT: TANG, Y. TOM
: APPLICANT: CORLEY, Neil C.
: APPLICANT: GUEGLER, Karl J.
: APPLICANT: BAUGHN, Mariah R.
: APPLICANT: LAL, Preeti
: APPLICANT: YUE, Henry
: APPLICANT: HILLMAN, Jennifer L.
: APPLICANT: AZIMZAI, Yalda
: TITLE OF INVENTION: HUMAN TRANSERASE PROTEINS
: FILE REFERENCE: PF-0592 PCT
: CURRENT APPLICATION NUMBER: US/09/786,240
: CURRENT FILING DATE: 2002-03-12
: PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 09/186,779; unassigned; 60/133,6
: PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1998-11-04; 1999-05-11
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: PERL Program
: SEQ ID NO 26
: LENGTH: 2065
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: Incyte ID No. 6558935 2617407CBI
US-09-786-240-26

```

Query Match	56.4%	Score 1153:	DB 4:	Length 2065:
Best Local Similarity	99.6%	Pred. No. 0:		
Matches 1593:	Conservative	0:	Mismatches	4: Indels
				2: Gaps
QY	106	CAAGTCTCCACTTCACACACATGCTACTGCTCTAAAAAATGAAAGCTCTGCTGTTCT	165	
DB	152	CAAGGTCTTCACCTTCACACACATGCTACTGCTCTAAAAAATGAAAGCTCTGCTGTTCT	211	
OY	166	GCTTTCACCAATGGCCATCTTGGCTCTATTCTCCACATGTATACCCCAACATCAGCTC	225	
DB	212	GCTTTCACCAATGGCCATCTTGGCTCTATTCTCCACATGTATACCCCAACATCAGCTC	271	
OY	226	CCTGTCTATGAAGGCACAGCCCGAGCGCATGCAGCTGCTGTCTGCTCTCTGCGCTC	285	
DB	272	CCTGTCTATGAAGGCACAGCCCGAGCGCATGCAGCTGCTGTCTGCTCTCTGCGCTC	331	
OY	286	TGCTCTCTCTTTTGTGGGGGACGCTTTTGGGACAGACCCACATGTTTCTTACTGATGGA	345	
DB	332	TGGCTCTCTCTTTTGTGGGGGACGCTTTTGGGACAGACCCACATGTTTCTTACTGATGGA	391	
OY	346	GCCCGCCTGGACAGTGTGATGACCTTCAACACAGACACCCGCTTGATGTGCACATGGC	405	
DB	392	GCCCGCCTGGACAGTGTGATGACCTTCAACACAGACACCCGCTTGATGTGCACATGGC	451	
OY	406	TGTGGGGATCTGATACGGGCCCTCTTTGTGCGACATGACCGCTTTGATGCCCTACAT	465	
DB	452	TGTGGGGATCTGATACGGGCCCTCTTTGTGCGACATGACCGCTTTGATGCCCTACAT	511	
OY	466	GGAACCTGTGTCGCGAGACAGTCAGGCTCTTTAGTGGGAGAAACAGCCGGGCCCTGTG	525	
DB	512	GGAACCTGTGTCGCGAGACAGTCAGGCTCTTTAGTGGGAGAAACAGCCGGGCCCTGTG	571	
OY	526	TTTGTGACCTGCTGTGACATCATCCCAAAAGATGAATTCATCCCGGGGCTCACTGCAG	585	
DB	572	TTTGTGACCTGCTGTGACATCATCCCAAAAGATG-AATTCATCCCGGGGCTCACTGCAG	630	
OY	586	GCTCCTGTGACATCAAGCCCTTTG-AAGTGTGAGAAAGAGCCGCGCTCTACAGCC	644	
DB	631	GCTCCTGTGACATCAAGCCCTTTGAAAGTTGTGGAGAAAGCCTGCGGCTCTACAGCC	690	
OY	645	ACGTGTGCTCAAGAGGATGCGCTTCTTCAACCTGACAGTCCCTTACCAGCTGTAAG	704	

Db	691	ACGTGTCCTCAAGAGAGGTGGCGTTCCTTCAACCTGCACTGTCAGTCCCTCTCAACCGGTGCTGAAG	750
QY	705	ACCCCTCCCTCAACCTGCATATTCGTGACACCTGTGTCGGGACCCCGGGCCGTGTTCCTT	764
Db	751	ACCCCTCCCTCAACCTGCATATTCGTGACACCTGTGTCGGGACCCCGGGCCGTGTTCCTT	810
QY	765	CCCGAAGACGACAAAGGAGATCTCATATGTTGACAGTGCAGATTGTGATGGGGCAGCATG	824
Db	811	CCCGAAGACGACAAAGGAGATCTCATATGTTGACAGTGCAGATTGTGATGGGGCAGCATG	870
QY	825	AGCAAAAATCTAAGAGAGGAGACCAACCTACTATGTGATGCAGTCACTGTGCCAAAGCC	884
Db	871	AGCAAAAATCTAAGAGAGGAGACCAACCTACTATGTGATGCAGTCACTGTGCCAAAGCC	930
QY	885	AGCGAGATCTACAAAGACCATCCAGTCCCTTGCCCAAGGCCCCGAGGAAGCGTACCTGCG	944
Db	931	AGCTGAGATCTACAAAGACCATCCAGTCCCTTGCCCAAGGCCCCGAGGAAGCGTACCTGCG	990
Y	945	TTGTGCGCTATGAGGACCTGGCTGAGGCCCTGTGAGCCAGACTTCCGGAATGTATGAT	1004
Db	991	TTGTGCGCTATGAGGACCTGGCTGAGGCCCTGTGAGCCAGACTTCCGGAATGTATGAT	1050
QY	1005	TCGTGGGATTTGGAAATTCCTTGCCCATCTTCCAGACTGGGTGCTAACTCAACCGAGCA	1066
Db	1051	TCGTGGGATTTGGAAATTCCTTGCCCATCTTCCAGACTGGGTGCTAACTCAACCGAGCA	1110
QY	1065	AGGCGATGGTGACACACGCTTTCACACAAATGCGAGGATGAGCCCTTAATGTCGCCAGG	1124
Db	1111	AGGCGATGGTGACACACGCTTTCACACAAATGCGAGGATGAGCCCTTAATGTCGCCAGG	1170
QY	1125	CTTGGCGCTGTCTTTCGCCATATGAAAAGTTTCTCGACTTCAGAAACCTGTGGCGATG	1184
Db	1171	CTTGGCGCTGTCTTTCGCCATATGAAAAGTTTCTCGACTTCAGAAACCTGTGGCGATG	1230
QY	1185	CCATGAATTTCTGGGCTACCGCCACGACATCTGAACAAAGAAACGAAACCGTGTGC	1244
Db	1231	CCATGAATTTCTGGGCTACCGCCACGACATCTGAACAAAGAAACGAAACCGTGTGC	1290
QY	1245	TGGATCTTCTGTCACTGGACTGTCCCTGAGCAAAATCACTAAGAGGGTTGAGAAAGCT	1304
Db	1291	TGGATCTTCTGTCACTGGACTGTCCCTGAGCAAAATCACTAAGAGGGTTGAGAAAGCT	1350
QY	1305	TTTGGCCACCTGGGTGACGCTCAGTCACTTCTCTGAATGCTTCTGAGCCTTGCCCTAC	1366
Db	1351	TTTGGCCACCTGGGTGACGCTCAGTCACTTCTCTGAATGCTTCTGAGCCTTGCCCTAC	1410
Y	1365	ATCTCTGAGCCTTAACATCATGTCTGTGGGTATCAACATGAGTGTGATTTGTGTCCACAC	1424
Db	1411	ATCTCTGAGCCTTAACATCATGTCTGTGGGTATCAACATGAGTGTGATTTGTGTCCACAC	1470
QY	1425	GTGCTCAAGCAAGAGACTTTTGTGTCCATGCTGTGTCTAGAAAACAGACTGGGGAAC	1484
Db	1471	GTGCTCAAGCAAGAGACTTTTGTGTCCATGCTGTGTCTAGAAAACAGACTGGGGAAC	1530
QY	1485	TTATGTGAGCAGACATCCCAACGATGAAGAAAGGGTATTTGCTCTTCTTTCTGTGATC	1544
Db	1531	TTATGTGAGCAGACATCCCAACGATGAAGAAAGGGTATTTGCTCTTCTTTCTGTGATC	1590
QY	1545	TTTCTGTCTGGGCAAGCTTTCAGAGACTTGTGGCCTGAGGCGCTATTAGCAGCAGACAG	1604
Db	1591	TTTCTGTCTGGGCAAGCTTTCAGAGACTTGTGGCCTGAGGCGCTATTAGCAGCAGACAG	1650
QY	1605	TATCAGTGGAAATTGATCCATAAACCTCCCTGTCCACATCTTGGCCAAATGGGGAATGATC	1664
Db	1651	TATCAGTGGAAATTGATCCATAAACCTCCCTGTCCACATCTTGGCCAAATGGGGAATGATC	1710
QY	1665	TTTCACCAAGAGGTCACACAGACTTTTCTCCAGAGATGCG	1703
Db	1711	TTTCACCAAGAGGTCACACAGACTTTTCTCCAGAGATGCG	1749

```

: Sequence 1 Application US/08992334
: Patent No. 5919678
: GENERAL INFORMATION:
: APPLICANT: Gruss, Alexandra
: APPLICANT: Maguin, Emmanuelle
: TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
: TITLE OF INVENTION: PLASMID
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Christie Parker & Hale, LLP
: STREET: 350 West Colorado Boulevard, Suite 500
: CITY: Pasadena
: STATE: California
: COUNTRY: United States
: ZIP: 91105
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/992.334
: FILING DATE: 17-DEC-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302.752
: FILING DATE: 24-DEC-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/FR93/00248
: FILING DATE: 12-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 992/03034
: FILING DATE: 13-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Prout, D. Bruce
: REGISTRATION NUMBER: 20958
: REFERENCE/DOCKET NUMBER: C93:31779
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (626) 795-9900
: TELEFAX: (626) 577-8800
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3792 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: circular
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: YES
: IMMEDIATE SOURCE:
: CLONE: pg-host4
:
: US-08-992-334-1
:
: Query Match 0.9%; Score 19; DB 2; Length 3792;
: Best Local Similarity 100.0%; Pred. No. 13;
: Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
:
: QY 1821 TCCTGACAAAGACGACG 1839
: Db 1818 TCCTGACAAAGACGACG 1836
:
: RESULT 5
: US-08-302-752-1
: Sequence 1, Application US/08302752
: Patent No. 6025190
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: THERMOSENSIBLE PLASMID
: NUMBER OF SEQUENCES: 3
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible

```


OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3792 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-1

Query Match 0.9%; Score 19; DB 3; Length 3792;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCCTGACAAAGAGCAAG 1839
|||||
DB 1818 TCCTGACAAAGAGCAAG 1836

RESULT 6
US-08-992-334-2
Sequence 2, Application US/08992334
Patent No. 5919678
GENERAL INFORMATION:
APPLICANT: Gruss, Alexandra
APPLICANT: Maguin, Emmanuelle
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christie Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR FR92/03034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 5234 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-992-334-2

Query Match 0.9%; Score 19; DB 2; Length 5234;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCCTGACAAAGAGCAAG 1839
|||||
DB 3260 TCCTGACAAAGAGCAAG 3278

RESULT 7
US-08-302-752-2
Sequence 2, Application US/08302752
Patent No. 6025190
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: THERMOSENSIBLE PLASMID
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5234 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-2

Query Match 0.9%; Score 19; DB 3; Length 5234;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCCTGACAAAGAGCAAG 1839
|||||
DB 3260 TCCTGACAAAGAGCAAG 3278

RESULT 8
US-08-992-334-3
Sequence 3, Application US/08992334
Patent No. 5919678
GENERAL INFORMATION:
APPLICANT: Gruss, Alexandra
APPLICANT: Maguin, Emmanuelle
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christie Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California

COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 932/03034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-8900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6722 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-992-334-3

Query Match
Best Local Similarity 100.0%; Score 19; DB 2; Length 6722;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCTCTGACAAAGAGCAAG 1839
|||||
DB 4748 TCTCTGACAAAGAGCAAG 4766

RESULT 9
S-08-302-752-3
Sequence 3, Application US/08302752
Patent No. 6025190
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: THERMOSENSIBLE PLASMID
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6722 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-3

Query Match
Best Local Similarity 100.0%; Score 19; DB 3; Length 6722;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCTCTGACAAAGAGCAAG 1839
|||||
DB 4748 TCTCTGACAAAGAGCAAG 4766

RESULT 10
US-09-020-956-69
Sequence 69, Application US/09020956
Patent No. 6261562
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, David C.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
NUMBER OF SEQUENCES: 178
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/020,956
FILING DATE: 09-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.427C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 536 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-09-020-956-69

Query Match
Best Local Similarity 100.0%; Score 18; DB 3; Length 536;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCATTGCTTGGG 18
|||||
DB 18 GAATTCATTGCTTGGG 35

RESULT 11
US-09-030-607-69
Sequence 69, Application US/09030607
Patent No. 6262245
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, David C.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS

NUMBER OF SEQUENCES: 224
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/030,607
FILING DATE: 25-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.427C3
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 536 base pairs
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-09-030-607-69

Query Match 0.9%; Score 18; DB 3; Length 536;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GAATTCATGTGTGGG 18
DB 18 GAATTCATGTGTGGG 35

RESULT 12
US-09-439-313-69
Sequence 69, Application US/09439313
Patent No. 6329505
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan Louise
APPLICANT: Jiang Yugu
APPLICANT: Reed, Steven G.
APPLICANT: Kalos, Michael
APPLICANT: Fanger, Gary
APPLICANT: Retter, Mark
APPLICANT: Solk, John
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
FILE REFERENCE: 210121.427C9
CURRENT APPLICATION NUMBER: US/09/439,313
NUMBER OF SEQ ID NOS: 575
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 69
LENGTH: 536
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc-feature

LOCATION: (1)...(536)
OTHER INFORMATION: n - A,T,C or G
US-09-439-313-69

Query Match 0.9%; Score 18; DB 4; Length 536;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GAATTCATGTGTGGG 18
DB 18 GAATTCATGTGTGGG 35

RESULT 13
US-09-352-616A-69
Sequence 69, Application US/09352616A
Patent No. 6395278
GENERAL INFORMATION:
APPLICANT: Dillon, Davin C.
APPLICANT: Harlocker, Susan Louise
APPLICANT: Jiang, Yugu
APPLICANT: Xu, Jiangchun
APPLICANT: Mitcham, Jennifer Lynn
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
FILE REFERENCE: 210121.427C8
CURRENT APPLICATION NUMBER: US/09/352,616A
CURRENT FILING DATE: 1999-07-13
NUMBER OF SEQ ID NOS: 472
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 69
LENGTH: 536
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc-feature
LOCATION: (1)...(536)
OTHER INFORMATION: n - A,T,C or G
US-09-352-616A-69

Query Match 0.9%; Score 18; DB 4; Length 536;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GAATTCATGTGTGGG 18
DB 18 GAATTCATGTGTGGG 35

RESULT 14
US-09-232-149A-69
Sequence 69, Application US/09232149A
Patent No. 6465611
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer Lynn
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
FILE REFERENCE: 210121.427C6
CURRENT APPLICATION NUMBER: US/09/232,149A
CURRENT FILING DATE: 1999-01-15
NUMBER OF SEQ ID NOS: 338
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 69
LENGTH: 536
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc-feature
LOCATION: (1)...(336)
OTHER INFORMATION: n - A,T,C or G
US-09-232-149A-69

Job time : 145 secs

Query Match 0.9%; Score 18; DB 4; Length 536;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 18 GAATTCATGTGTGGG 35

RESULT 15

US-09-107-532A-1333
; Sequence 1333 Application US/09107532A
; Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariunello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 1333:

SEQUENCE CHARACTERISTICS:

LENGTH: 699 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:

NAME/KEY: misc_feature

LOCATION: (B) LOCATION 1...699

SEQUENCE DESCRIPTION: SEQ ID NO: 1333:

US-09-107-532A-1333

Query Match 0.9%; Score 18; DB 4; Length 699;
Best Local Similarity 100.0%; Pred. No. 41;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1528 TTCTTCTTTCTTGATCT 1545
|||||
DB 473 TTCTTCTTTCTTGATCT 490

Search completed: August 15, 2003, 08:09:08

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2003, 04:16:14 ; Search time 537 Seconds
(without alignments)
8514.175 Million cell updates/sec

Title: US-09-645-078-1
Perfect score: 2043
Sequence: 1 gattccattgtgtgtgta.....tgggacccaataaaaaa 2043

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 1504479 seqs, 1118970152 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1763610

Minimum DB seq length: 300
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications_NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1547	75.7	2032	9	US-09-816-825-1
2	1547	75.7	2032	13	US-10-007-262-2
3	403	19.7	505	10	US-09-898-598-2595
4	403	19.7	517	10	US-09-919-580-194
5	114	5.6	389	9	US-09-919-580-264
6	59	2.9	2544	9	US-09-927-602-1
7	59	2.9	48436	9	US-09-927-602-38
8	20	1.0	398	10	US-09-960-352-11099
9	20	1.0	570	13	US-10-027-632-140660
10	20	1.0	11102	14	US-10-205-823-334
11	20	1.0	180557	13	US-10-003-806-6
12	20	0.9	180557	13	US-10-003-806-9
13	19	0.9	341	10	US-09-867-701-8861
14	19	0.9	341	10	US-09-867-701-8868
15	19	0.9	349	9	US-09-759-143-649
16	19	0.9	349	9	US-09-780-669-649

17	19	0.9	349	9	US-09-822-827-649	Sequence 649, App
18	19	0.9	349	10	US-09-895-793-649	Sequence 649, App
19	19	0.9	349	10	US-09-895-814-649	Sequence 649, App
20	19	0.9	349	13	US-10-012-896-649	Sequence 649, App
21	19	0.9	405	10	US-09-867-701-9634	Sequence 9634, App
22	19	0.9	544	13	US-10-027-632-87347	Sequence 87347, A
23	19	0.9	786	13	US-10-027-632-118925	Sequence 118925,
24	19	0.9	2874	13	US-10-027-632-111785	Sequence 111785,
25	19	0.9	3273	12	US-09-971-392-26	Sequence 26, Appl
26	19	0.9	3274	13	US-10-002-600-52	Sequence 52, Appl
27	19	0.9	4224	9	US-09-815-242-9944	Sequence 9944, Ap
28	19	0.9	4537	13	US-10-108-605-322	Sequence 322, App
29	19	0.9	6546	12	US-10-017-161-2201	Sequence 2201, Ap
30	19	0.9	8742	9	US-09-908-711-147	Sequence 147, App
31	19	0.9	8742	11	US-09-764-891-6217	Sequence 6217, Ap
32	19	0.9	10929	14	US-10-032-393-7	Sequence 7, Appl
33	19	0.9	62944	10	US-09-954-456-2257	Sequence 2257, Ap
34	19	0.9	75899	9	US-09-854-883-243	Sequence 243, App
35	19	0.9	684973	10	US-09-263-959-1	Sequence 1, Appl
36	18	0.9	300	9	US-09-767-870-7	Sequence 7, Appl
37	18	0.9	300	12	US-10-242-568-7	Sequence 7, Appl
38	18	0.9	303	10	US-09-867-701-9733	Sequence 9733, Ap
39	18	0.9	309	12	US-09-237-183A-2108	Sequence 2108, Ap
40	18	0.9	339	10	US-09-867-701-9013	Sequence 9013, Ap
41	18	0.9	362	10	US-09-867-701-9419	Sequence 9419, Ap
42	18	0.9	389	10	US-09-867-701-9251	Sequence 9251, Ap
43	18	0.9	389	10	US-09-867-701-9930	Sequence 9930, Ap
44	18	0.9	391	10	US-09-867-701-9163	Sequence 9163, Ap
45	18	0.9	394	10	US-09-867-701-8735	Sequence 8735, Ap

ALIGNMENTS

RESULT 1
US-09-816-825-1
Sequence 1, Application us/09816825
Patent No. US20010051370A1
GENERAL INFORMATION:
APPLICANT: Bistrup, Annette
APPLICANT: Rosen, Steven D.
APPLICANT: Hemmerich, Stefan
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
FILE REFERENCE: 6510-107CON
CURRENT APPLICATION NUMBER: US/09/816, 825
CURRENT FILING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: 09/045, 284
PRIOR FILING DATE: 1998-03-20
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2032
TYPE: DNA
ORGANISM: Homo sapiens
US-09-816-825-1

Query Match 75.7% Score 1547; DB 9; Length 2032;
Best Local Similarity 99.9% Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 106 CAAGGTCCTTCACCTTCAGCACAATGCTACTGCTTAATAAATGAAGTCTGCTGTTCT 165
DB 160 CAAGGTCCTTCACCTTCAGCACAATGCTACTGCTTAATAAATGAAGTCTGCTGTTCT 219
QY 166 GATTTCCTTCAGATGCGCATCTTGGCTCTATCTTCACATGTACAGCCACAATCAGCTC 225
DB 220 GATTTCCTTCAGATGCGCATCTTGGCTCTATCTTCACATGTACAGCCACAATCAGCTC 279
QY 226 CCGTGTCTTAAAGGACAGCCGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 285
DB 280 CCGTGTCTTAAAGGACAGCCGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 339
QY 286 TGCGTCTTCTTTTGTGGGCGAGCTTTTGGGCGAGCACCAGATGTTTCTACTGATGGA 345

Db 340 TGGCTCTCTTTTGGGGGACCTTTTGGGAGCACCAGATGTTTCTACCTGATGGA 399
 QY 346 GCGCGCTGGACAGTGTGATGACCTTCAAGCAGAGACACCGCTGGATCTGTGACATGGC 405
 Db 400 GCGCGCTGGACAGTGTGATGACCTTCAAGCAGAGACACCGCTGGATCTGTGACATGGC 459
 QY 406 TGTGGGGGATCTGATACGGGGCGTCTTGTGTGACATGAGCGTCTTGTGATGCTTACAT 465
 Db 460 TGTGGGGGATCTGATACGGGGCGTCTTGTGTGACATGAGCGTCTTGTGATGCTTACAT 519
 QY 466 GGAACCTGGTCCCGGAGACATCCAGCTCTTCACTGAGGAGAACAGCGCGCGCTGTG 525
 Db 520 GGAACCTGGTCCCGGAGACATCCAGCTCTTCACTGAGGAGAACAGCGCGCGCTGTG 579
 QY 526 TTCTGACCTGCTGTGACATCATCCACAAATGATGATATCCCGGGCTCACTGACG 585
 Db 580 TTCTGACCTGCTGTGACATCATCCACAAATGATGATATCCCGGGCTCACTGACG 639
 QY 586 GCTCTGTGACATCAACAGCCCTTGTGAGGTGTGAGAGAGCGCTGCTGTGACAGCA 645
 Db 640 GCTCTGTGACATCAACAGCCCTTGTGAGGTGTGAGAGAGCGCTGCTGTGACAGCA 699
 QY 646 GCTGTGCTCAAGAGAGGTGCTGTCTCAACCTGAGTCCCTTACCTGCTGTGAGGA 705
 Db 700 GCTGTGCTCAAGAGAGGTGCTGTCTTCAACCTGAGTCCCTTACCTGCTGTGAGGA 759
 QY 706 CCGCTCCCTCAACCTGATATGTCACACCTGCTGCGGAGCCCGGCGCTGTGCTGTC 765
 Db 760 CCGCTCCCTCAACCTGATATGTCACACCTGCTGCGGAGCCCGGCGCTGTGCTGTC 819
 QY 766 CCGAGAGCGCAAAAGGAGATCTCATGTTGACAGTGTGATGAGTGTGAGGAGAGATCA 825
 Db 820 CCGAGAGCGCAAAAGGAGATCTCATGTTGACAGTGTGATGAGTGTGAGGAGAGATCA 879
 QY 826 GCAAAACTCAAGAGAGAGACCAACCTACTATGATGAGAGATGATGCAAGCA 885
 Db 880 GCAAAACTCAAGAGAGAGACCAACCTACTATGATGAGAGATGATGCAAGCA 939
 QY 886 GCTGAGATCTCAAGAGACATCAAGTCTTGGCCCAAGCGCTGTGAGAGAGATGCT 945
 Db 940 GCTGAGATCTCAAGAGACATCAAGTCTTGGCCCAAGCGCTGTGAGAGAGATGCT 999
 QY 946 TGTGGCTGTGAGAGAGCTGGCTGAGCGCCCTGAGCGCCAGATCCCGGAATGTATGAAT 1005
 Db 1000 TGTGGCTGTGAGAGAGCTGGCTGAGCGCCCTGAGCGCCAGATCCCGGAATGTATGAAT 1059
 QY 1006 CGTGGATTTGGAATCTTGGCCCATCTTCAAGCTGTGATGATTAACATCAGCGAGCA 1065
 Db 1060 CGTGGATTTGGAATCTTGGCCCATCTTCAAGCTGTGATGATTAACATCAGCGAGCA 1119
 QY 1066 GGGCATGGTGAACAGCTTTCACACAAATGCGAGGATGCCCTTAATGTCTCCAGGC 1125
 Db 1120 GGGCATGGTGAACAGCTTTCACACAAATGCGAGGATGCCCTTAATGTCTCCAGGC 1179
 QY 1126 TTTGGGCTGTGAGAGAGCTGGCTGAGCGCCCTGAGCGCCAGATCCCGGAATGTATGAAT 1185
 Db 1180 TTTGGGCTGTGAGAGAGCTGGCTGAGCGCCCTGAGCGCCAGATCCCGGAATGTATGAAT 1239
 QY 1186 CATGAATTTGCTGGGCTACCGCAGCTGATCTGAACAAAGAGAGAACTGTTGCT 1245
 Db 1240 CATGAATTTGCTGGGCTACCGCAGCTGATCTGAACAAAGAGAGAACTGTTGCT 1299
 QY 1246 GGAATCTTGTCTACCTGAGCTGTCCCTGAGCAATCCACTAAGAGGTTGAGAAAGCTT 1305
 Db 1300 GGAATCTTGTCTACCTGAGCTGTCCCTGAGCAATCCACTAAGAGGTTGAGAAAGCTT 1359
 QY 1306 TGTGAGAGCTGTGAGAGCTGAGTCACTTCTGATGATGCTTGTGAGAGCTGTGATG 1365
 Db 1360 TGTGAGAGCTGTGAGAGCTGAGTCACTTCTGATGATGCTTGTGAGAGCTGTGATG 1419
 QY 1366 TGTGAGAGCTTACTACATGCTGTGGTATCAACTGATGATGATGATGATGATGATG 1425

Db 1420 TGTGAGAGCTTACTACATGCTGTGGTATCACTGAGTGTGATGATGATGATGATG 1479
 QY 1426 TGTGAGAGAGAGAGAGCTTTGTGTCATGCTGTGTGATGAGAAACAGAGAGAGAGAGCT 1485
 Db 1480 TGTGAGAGAGAGAGAGCTTTGTGTCATGCTGTGTGATGAGAAACAGAGAGAGAGAGCT 1539
 QY 1486 TATGTGAGAGAGAGAGAGCTTGTGTCATGCTGTGTGATGAGAAACAGAGAGAGAGAGCT 1545
 Db 1540 TATGTGAGAGAGAGAGAGCTTGTGTCATGCTGTGTGATGAGAAACAGAGAGAGAGAGCT 1599
 QY 1546 TGTGAGAGAGAGAGAGCTTGTGTCATGCTGTGTGATGAGAAACAGAGAGAGAGAGCT 1605
 Db 1600 TGTGAGAGAGAGAGAGCTTGTGTCATGCTGTGTGATGAGAAACAGAGAGAGAGAGCT 1659
 QY 1606 ATCAGTGAATGATGATCAATTAACCTCCCTGTGATGATGATGATGATGATGATGATG 1665
 Db 1660 ATCAGTGAATGATGATCAATTAACCTCCCTGTGATGATGATGATGATGATGATGATG 1719
 QY 1666 TTACCAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
 Db 1720 TTACCAAGAGCTCACAGCATTTTCCACAGAGATGC 1757
 RESULT 2
 US-10-007-262-2
 ; Sequence 2, Application US/10007262
 ; Publication No. US20020164748A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bistrup, Annette
 ; APPLICANT: Rosen, Steven D.
 ; APPLICANT: Tangemann, Kirsten
 ; APPLICANT: Hemmerich, Stefan
 ; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
 ; FILE REFERENCE: 6510-107CIP
 ; CURRENT APPLICATION NUMBER: US/10/007,262
 ; PRIOR FILING DATE: 2001-11-08
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 2032
 ; TYPE: DNA
 ; ORGANISM: H. sapiens
 ; US-10-007-262-2
 Query Match 75.7%; Score 1547; DB 13; Length 2032;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 106 CAAGTCTTCCACTTCAGCAATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTCT 165
 Db 160 CAAGTCTTCCACTTCAGCAATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTCT 219
 QY 166 GGTTCGCCAGATGGGCATTTGGCTCTATTTCCACATGTAAGGACCAATCATGAGCTC 225
 Db 220 GGTTCGCCAGATGGGCATTTGGCTCTATTTCCACATGTAAGGACCAATCATGAGCTC 279
 QY 226 CCTGTATGAAGGACAGACCCGAGCGCATGACAGTGTGCTGTCTTCTGCGCTC 285
 Db 280 CCTGTATGAAGGACAGACCCGAGCGCATGACAGTGTGCTGTCTTCTGCGCTC 339
 QY 286 TGGCTCTTTTGTGGGGACACTTTTGGGAGACCAAGATGTTTCTACTGATGA 345
 Db 340 TGGCTCTTTTGTGGGGACACTTTTGGGAGACCAAGATGTTTCTACTGATGA 399
 QY 346 GCGCGCTGGACAGTGTGATGACCTTCAAGCAGAGACACCGCTGGATGCTGACATGGC 405
 Db 400 GCGCGCTGGACAGTGTGATGACCTTCAAGCAGAGACACCGCTGGATGCTGACATGGC 459
 QY 406 TGTGAGAGATCTGATACGGGGCGTCTTGTGTGACATGAGCGTCTTGTGATGCTTACAT 465
 Db 460 TGTGAGAGATCTGATACGGGGCGTCTTGTGTGACATGAGCGTCTTGTGATGCTTACAT 519

QY 466 GGAACCTGTGTCGGGAGAGAGTCACGCTCTTTCAGTGGGAGAGACCGGCGCTGTG 525
 DB 520 GGAACCTGTGTCGGGAGAGAGTCACGCTCTTTCAGTGGGAGAGACCGGCGCTGTG 579
 QY 526 TTCTGACGCTGCTGTGATCATCCACAGATGAATATCCCGGGGCTCACTGAG 585
 DB 580 TTCTGACGCTGCTGTGATCATCCACAGATGAATATCCCGGGGCTCACTGAG 639
 QY 586 GCTCCCTGTGAGTCAAGAGCCCTTGTAGGTGTGGAGAGAGCCCTGCTCTACAGCA 645
 DB 640 GCTCCCTGTGAGTCAAGAGCCCTTGTAGGTGTGGAGAGAGCCCTGCTCTACAGCA 699
 QY 646 CGTGTGCTCAAGAGAGTGGGCTCTTCAACCTGAGTCCCTTACCCGGCTGAGAAAGA 705
 DB 700 CGTGTGCTCAAGAGAGTGGGCTCTTCAACCTGAGTCCCTTACCCGGCTGAGAAAGA 759
 QY 706 CCCCTCCCTCAACCTGATATCGTACACCTGGTCCGGGACCCCGGGCGGTTCGTTTC 765
 DB 760 CCCCTCCCTCAACCTGATATCGTACACCTGGTCCGGGACCCCGGGCGGTTCGTTTC 819
 QY 766 CCGAGAACGACAAAGGAGATCTCATGATTTGACAGTGCATTTGTATGGGGGACATGA 825
 DB 820 CCGAGAACGACAAAGGAGATCTCATGATTTGACAGTGCATTTGTATGGGGGACATGA 879
 QY 826 GCAAAAACCTCAAGAGAGAGACCAACCTTACTATGTATGATGAGGTGATCTGCCAAAGCA 885
 DB 880 GCAAAAACCTCAAGAGAGAGACCAACCTTACTATGTATGATGAGGTGATCTGCCAAAGCA 939
 QY 886 GCTGTGATCTACAAAGACCATCCAGTCTTGGCCCAAGCCCTGACAGAACGCTACTGCT 945
 DB 940 GCTGTGATCTACAAAGACCATCCAGTCTTGGCCCAAGCCCTGACAGAACGCTACTGCT 999
 QY 946 TTGTCCCTATGAGAGAGTGGCTGTGAGCCCTGTGGCCCAACTTCCGATGTATGATTT 1005
 DB 1000 TTGTCCCTATGAGAGAGTGGCTGTGAGCCCTGTGGCCCAACTTCCGATGTATGATTT 1059
 QY 1006 CGTGTGATGAGAGAGTGGCTGTGAGCCCTGTGGCCCAACTTCCGATGTATGATTT 1065
 DB 1060 CGTGTGATGAGAGAGTGGCTGTGAGCCCTGTGGCCCAACTTCCGATGTATGATTT 1119
 QY 1066 GGGGATGGTACACACGCTTTCACACAAATGCCAGGATGCCCTTATGTCTCCAGGC 1125
 DB 1120 GGGGATGGTACACACGCTTTCACACAAATGCCAGGATGCCCTTATGTCTCCAGGC 1179
 QY 1126 TTGGCGCTGTGCTTGGCCCTATGAAAAGTTTCTGAGCTTCAAGAAACCTGTGGCATTC 1185
 DB 1180 TTGGCGCTGTGCTTGGCCCTATGAAAAGTTTCTGAGCTTCAAGAAACCTGTGGCATTC 1239
 QY 1186 CATGAATTTGCTGGGCTACCGCCACGTCAGATCTGAACAAGAGAAACCTGTGCT 1245
 DB 1240 CATGAATTTGCTGGGCTACCGCCACGTCAGATCTGAACAAGAGAAACCTGTGCT 1299
 QY 1246 GGATCTTCTGTACTGAGCTGTCCCTGAGCAATCCACTAAGAGGTTGAGAAAGCTT 1305
 DB 1300 GGATCTTCTGTACTGAGCTGTCCCTGAGCAATCCACTAAGAGGTTGAGAAAGCTT 1359
 QY 1306 TGTGCGACCTGTGTAGAGCTGATGATCTTCTGTAATGCTTCAAGCTTGTCCATCA 1365
 DB 1360 TGTGCGACCTGTGTAGAGCTGATGATCTTCTGTAATGCTTCAAGCTTGTCCATCA 1419
 QY 1366 TCTGTAGAGCTTAACTACATGTCTGTGGTATCAACATGAGTGTGATGTGTCACAGC 1425
 DB 1420 TCTGTAGAGCTTAACTACATGTCTGTGGTATCAACATGAGTGTGATGTGTCACAGC 1479
 QY 1426 TGTCTCAAGCAGAGAGACTTTTGTGTGTCATGCTGTGTAAGAAACAGACTGGGAACT 1485
 DB 1480 TGTCTCAAGCAGAGAGACTTTTGTGTGTCATGCTGTGTAAGAAACAGACTGGGAACT 1539
 QY 1486 TATGTGAGCAGACATCCACAGTGAAGAGGATATGCTCTTCTTCTTGTGATCT 1545
 DB 1540 TATGTGAGCAGACATCCACAGTGAAGAGGATATGCTCTTCTTCTTGTGATCT 1599

QY 1546 TCCTGTGCGGACACTTCAGAGACTTGTGGCTGTGAGAGCCATTAAGACAGACAGT 1605
 DB 1600 TCCTGTGCGGACACTTCAGAGACTTGTGGCTGTGAGAGCCATTAAGACAGACAGT 1659
 QY 1606 ATCAGTGAATTTGATCCATAAATCCCTGTGCATCTTTGCCAATGGGAAATGATCT 1665
 DB 1660 ATCAGTGAATTTGATCCATAAATCCCTGTGCATCTTTGCCAATGGGAAATGATCT 1719
 QY 1666 TTCACCAAGAGCTCAGCAGATTTCCACAGAGATGC 1703
 DB 1720 TTCACCAAGAGCTCAGCAGATTTCCACAGAGATGC 1757

RESULT 3
 US-09-998-598-2595
 : Sequence 2595, Application US/09998598
 : Patent No. US20020150922A1
 : GENERAL INFORMATION:
 : APPLICANT: Stolk, John A.
 : APPLICANT: Xu, Jiangchun
 : APPLICANT: Chenault, Ruth A.
 : APPLICANT: Mesinger, Madelein Joy
 : TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 : FILE REFERENCE: 210121.561
 : CURRENT APPLICATION NUMBER: US/09/998,598
 : NUMBER OF SEQ ID NOS: 2606
 : SOFTWARE: Corixa Invention Disclosure Database
 : SEQ ID NO 2595
 : LENGTH: 505
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 US-09-998-598-2595

Query Match 19.7%; Score 403; DB 10; Length 505;
 Best Local Similarity 100.0%; Pred. No. 8.5e-203;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GGCCTTGTGCTGACCTGCTGTGACGCTCAGCTCACTTTCTCTGAATGCTTGTGACCTTGC 1360
 DB 1 GGCCTTGTGCTGACCTGCTGTGACGCTCAGCTCACTTTCTCTGAATGCTTGTGACCTTGC 60
 QY 1361 CTACATCTGAGGCTTAACATGATGCTGTGGTATCACAGTGAAGTGTGATGCTG 1420
 DB 61 CTACATCTGAGGCTTAACATGATGCTGTGGTATCACAGTGAAGTGTGATGCTG 120
 QY 1421 ACAGTGTCTCAAGCAGAGAGACTTTGTGTCCATGCTTGTGTCTAGAAAACAGACTGGG 1480
 DB 121 ACAGTGTCTCAAGCAGAGAGACTTTGTGTCCATGCTTGTGTCTAGAAAACAGACTGGG 180
 QY 1481 AACCTTATGTGAGCAGACATCCACAGTGAAGAGGATTTGCTTCTTCTTCTTCTT 1540
 DB 181 AACCTTATGTGAGCAGACATCCACAGTGAAGAGGATTTGCTTCTTCTTCTTCTT 240
 QY 1541 GATTTCTGTGCTGGGAGACTTGTGAGACTTGTGGCTGGAGAGCCATTAAGCAGAC 1600
 DB 241 GATTTCTGTGCTGGGAGACTTGTGAGACTTGTGGCTGGAGAGCCATTAAGCAGAC 300
 QY 1601 ACATATCAGTGAATTTGATCCATAAATCCCTGTGCATCTTGTGCCAATGGGAGATG 1660
 DB 301 ACATATCAGTGAATTTGATCCATAAATCCCTGTGCATCTTGTGCCAATGGGAGATG 360
 QY 1661 GATTTTACCAAGAGACTCAGCAGCATTTTTCACAGAGATGC 1703
 DB 361 GATTTTACCAAGAGACTCAGCAGCATTTTTCACAGAGATGC 403

RESULT 4
 US-09-919-580-194
 : Sequence 194, Application US/09919580
 : Patent No. US20020110832A1
 : GENERAL INFORMATION:

```

; APPLICANT: Pyle, Ruth
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secret, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.552
; CURRENT APPLICATION NUMBER: US/09/919,580
; CURRENT FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 194
; LENGTH: 517
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-919-580-194

```

```

Query Match          19.7%; Score 403; DB 10; Length 517;
Best Local Similarity 100.0%; Pred. No. 8.5e-203; Indels 0; Gaps 0;
Matches 403; Conservative 0; Mismatches 0;

```

```

QY 1301 GCGTTTGCCACCTGCTGAGCCTGCTGAGCTTCTGTAATGCTTCTGAGCCTTGC 1360
DB 13 GCGTTTGCCACCTGCTGAGCCTGCTGAGCTTCTGTAATGCTTCTGAGCCTTGC 72
QY 1361 CTACATCTGAGCCTTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1420
DB 73 CTACATCTGAGCCTTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 132
QY 1421 ACACGTGCTCAAGCAGAGAGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1480
DB 133 ACACGTGCTCAAGCAGAGAGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 192
QY 1481 AACCTTATGTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1540
DB 193 AACCTTATGTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 252
QY 1541 GATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1600
DB 253 GATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 312
QY 1601 ACAGTATCAGTGAATGATTCATTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1660
DB 313 ACAGTATCAGTGAATGATTCATTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 372
QY 1661 GATCTTCAACCAAGAGCTCAGCAGCAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1703
b 373 GATCTTCAACCAAGAGCTCAGCAGCAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 415

```

RESULT 5

```

US-09-919-580-264
; Sequence 264, Application US/09919580
; Patent No. US20020110832A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secret, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.552
; CURRENT APPLICATION NUMBER: US/09/919,580
; CURRENT FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 264
; LENGTH: 389
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-919-580-264

```

```

Query Match          5.6%; Score 114; DB 10; Length 389;
Best Local Similarity 100.0%; Pred. No. 1.3e-49;
Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1930 TGCCATCTGTATATACTAAATCCCAATTAAGTTCTGTAGATGTCCCTTTTATG 1989
DB 266 TGCCATCTGTATATACTAAATCCCAATTAAGTTCTGTAGATGTCCCTTTTATG 325
QY 1990 CTCTTAATTAATAGCAGTAATGTCATTTTATGAGATCTTAATTAATTAATTAAT 2043
DB 326 CTCTTAATTAATAGCAGTAATGTCATTTTATGAGATCTTAATTAATTAATTAAT 379

```

RESULT 6

```

US-09-927-602-1
; Sequence 1, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2544
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (693)...(1877)
US-09-927-602-1

```

```

Query Match          2.9%; Score 59; DB 9; Length 2544;
Best Local Similarity 100.0%; Pred. No. 1.9e-20; Indels 0; Gaps 0;
Matches 59; Conservative 0; Mismatches 0;

```

```

QY 624 AGGCTGCGCGCTCTCTACAGCAGCAGTGCTCAAGAGAGTGCGCTTCTCAACCTGAG 682
DB 1180 AGGCTGCGCGCTCTCTACAGCAGCAGTGCTCAAGAGAGTGCGCTTCTCAACCTGAG 1238

```

RESULT 7

```

US-09-927-602-38
; Sequence 38, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 48436
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-927-602-38

```

```

Query Match          2.9%; Score 59; DB 9; Length 48436;
Best Local Similarity 100.0%; Pred. No. 1.9e-20; Indels 0; Gaps 0;
Matches 59; Conservative 0; Mismatches 0;

```

```

QY 624 AGGCTGCGCGCTCTCTACAGCAGCAGTGCTCAAGAGAGTGCGCTTCTCAACCTGAG 682
DB 48024 AGGCTGCGCGCTCTCTACAGCAGCAGTGCTCTCAAGAGAGTGCGCTTCTCAACCTGAG 48082

```


RESULT 8
US-09-960-352-11099/C
; Sequence 11099, Application US/09960352
; Patent No. US2002013719A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengping
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalaagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; LENGTH: 398
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 48-BOVMS1-008-Q1-E1-DB
US-09-960-352-11099

Query Match 1.0%; Score 20; DB 10; Length 398;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 175 GATGCCATCTTGGCTCTAT 194
DB 240 GATGCCATCTTGGCTCTAT 221

RESULT 9
US-10-027-632-140660
; Sequence 140660, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 140660
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-140660

Query Match 1.0%; Score 20; DB 13; Length 570;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 700 GAAGACCCCTCCCAACC 719
DB 273 GAAGACCCCTCCCAACC 292

RESULT 10
US-10-205-823-334/C
; Sequence 334, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhang
; APPLICANT: Monsey, Angela M.
; APPLICANT: Glatz, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 334
; LENGTH: 11102
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-205-823-334

Query Match 1.0%; Score 20; DB 14; Length 11102;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1523 TGCTCTCTCTCTTCTTGA 1542
DB 722 TGCTCTCTCTCTTCTTGA 703

RESULT 11
US-10-003-806-6
; Sequence 6, Application US/10003806
; Publication No. US20020119929A1
; GENERAL INFORMATION:
; APPLICANT: Bishop, Colin E.
; APPLICANT: Agoulunk, Alexander I.
; APPLICANT: Zhu, Qichao
; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
; FILE REFERENCE: P020660S1/10024824
; CURRENT APPLICATION NUMBER: US/10/003,806
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 60/245,872
; PRIOR FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 180557
; TYPE: DNA
; ORGANISM: Human
US-10-003-806-6

Query Match 1.0%; Score 20; DB 13; Length 180557;

Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1893 TTTTCCATTACATAGAAA 1912
|||||
DB 2567 TTTTCCATTACATAGAAA 2586

RESULT 12

US-10-003-806-9
; Sequence 9, Application US/10003806
; Publication No. US20020119929A1
; GENERAL INFORMATION:

APPLICANT: Bishop, Colin E.
APPLICANT: Agoulnik, Alexander I.
APPLICANT: Zhu, Qichao
TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
FILE REFERENCE: P02066US1/10024824
CURRENT APPLICATION NUMBER: US/10/003,806
CURRENT FILING DATE: 2001-11-02
PRIOR APPLICATION NUMBER: 60/245,872
PRIOR FILING DATE: 2000-11-03
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 9
LENGTH: 180557
TYPE: DNA
ORGANISM: Human
US-10-003-806-9

Query Match

Best Local Similarity 1.0%; Score 20; DB 13; Length 180557;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1893 TTTTCCATTACATAGAAA 1912
|||||
DB 2567 TTTTCCATTACATAGAAA 2586

RESULT 13

US-09-867-701-8841/C
; Sequence 8841, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:

APPLICANT: Aglate, Paul A.
APPLICANT: Jones, Robert
APPLICANT: Harlocker, Susan L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.497
CURRENT APPLICATION NUMBER: US/09/867,701
CURRENT FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 10912
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8841
LENGTH: 341
TYPE: DNA
ORGANISM: Homo sapien
US-09-867-701-8841

Query Match

Best Local Similarity 0.9%; Score 19; DB 10; Length 341;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAAGACAGACAGAGGTAG 41
|||||
DB 330 GAAGACAGACAGAGGTAG 312

RESULT 14

US-09-867-701-8888/C
; Sequence 8888, Application US/09867701
; Patent No. US20020132237A1

GENERAL INFORMATION:

APPLICANT: Aglate, Paul A.
APPLICANT: Jones, Robert
APPLICANT: Harlocker, Susan L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.497
CURRENT APPLICATION NUMBER: US/09/867,701
CURRENT FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 10912
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8888
LENGTH: 341
TYPE: DNA
ORGANISM: Homo sapien
US-09-867-701-8888

Query Match

Best Local Similarity 0.9%; Score 19; DB 10; Length 341;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAAGACAGACAGAGGTAG 41
|||||
DB 330 GAAGACAGACAGAGGTAG 312

RESULT 15

US-09-759-143-649
; Sequence 649, Application US/09759143
; Patent No. US2002022248A1
; GENERAL INFORMATION:

APPLICANT: Xu, Jianshun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqun
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skelky, Yasir A.W.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.427C23
CURRENT APPLICATION NUMBER: US/09/759,143
CURRENT FILING DATE: 2001-01-12
NUMBER OF SEQ ID NOS: 934
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 649
LENGTH: 349
TYPE: DNA
ORGANISM: Homo sapien

Query Match

Best Local Similarity 0.9%; Score 19; DB 9; Length 349;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCATTGTGTGGGT 19
|||||
DB 72 GAATTCATTGTGTGGGT 90

Fri Aug 15 08:09:28 2003

us-09-645-078-1.rnpb

Page 7

Search completed: August 15, 2003, 08:06:32
Job time : 539 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:38:24 ; Search time 39.3979 Seconds

(without alignments)
1555.119 Million cell updates/sec

Title: US-09-645-078-2

Perfect score: 2038

Sequence: 1 MLPPKKMLLFLVQMAIL.....EQRNLDDLSTWTPQIH 386

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	2038	100.0	386 20	AAV39918 Human glycosyl sul
2	2035	99.9	386 22	AAV3309 Human polypeptide,
3	2008	98.5	380 23	AAU11274 Human L-selectin s
4	1936	95.0	386 21	AAV79219 Human transferrase
5	1513.5	74.3	388 20	AAV39919 Mouse glycosyl sul
6	1205.5	59.2	418 21	AAV1947 Human ORFX ORF1711
7	1028	50.4	418 23	ABB1557 Mouse intestinal N
8	1019	50.0	395 22	AAV72640 Human glycosyl sul
9	1019	50.0	395 23	ABB1554 Human corneal N-ac

10	1019	50.0	395 23	AAE15438	Human drug metabol
11	1017.5	49.9	395 22	AAV72638	Mouse glycosyl sul
12	1017.5	49.9	395 23	AAU11275	Murine intestinal-
13	1013.5	49.7	395 23	ABB1555	Consensus N-acetyl
14	1008	49.5	390 22	AAV72639	Human glycosyl sul
15	1008	49.5	390 23	ABB1556	Human intestinal N
16	821	40.3	171 23	ABB1560	Human high endothe
17	605.5	29.7	486 24	ABP56121	Human chondroitin
18	598.5	29.4	483 20	AAV31656	Mouse N-acetylgluc
19	591.5	29.0	530 22	AAV39367	Human protein sequ
20	587.5	28.8	484 20	AAV31657	Human N-acetylgluc
21	587.5	28.8	531 23	AAU69414	Lung small cell ca
22	549	26.9	411 19	AAV61100	Keratan sulphate 6
23	549	26.9	411 24	AAE23356	Human chondroitin
24	549	26.9	411 24	ABU03503	Angiogenesis-assoc
25	519.5	25.5	169 23	ABB1558	Human corneal N-ac
26	511.5	25.1	169 23	ABB1559	Human intestinal N
27	500.5	24.6	458 18	AAW06480	Chick chondroitin
28	492	24.1	499 24	ABR41139	Human DTRP enzyme
29	482	23.7	479 19	AAW52863	Glycosaminoglycan
30	336	16.5	169 23	ABB1561	Human N-acetylgluc
31	335.5	16.5	179 23	ABB1562	Human keratan sulf
32	293	14.4	174 23	ABB1563	Human chondroitin-
33	210	10.3	363 22	ABB64512	Drosophila melanog
34	186.5	9.2	183 22	ABB68582	Mycobacterium tube
35	116	5.7	388 24	AAE32782	Mouse glycosyl sul
36	109	5.3	1207 22	AAV72643	Human novel Cpg-as
37	109	5.3	1207 24	AAE33542	Human novel Cpg-as
38	108.5	5.3	596 22	AAV72641	Human novel Cpg-as
39	108.5	5.3	1212 24	AAE33541	Human glycosyl sul
40	108.5	5.3	1222 22	AAV72642	Human secreted pro
41	108.5	5.3	1222 24	ABU11849	Human protein sequ
42	100.5	4.9	775 23	ABB97942	Human nucleic acid
43	96.5	4.7	743 22	ABU52986	Human transporter
44	96.5	4.7	791 23	AAE22917	Human NOV16a, tumo
45	95.5	4.7	568 23	ABB98418	

ALIGNMENTS

RESULT 1	
AAV39918	AAV39918 standard; Protein: 386 AA.
ID	AAV39918;
AC	AAV39918;
XX	08-DEC-1999 (first entry)
DT	Human glycosyl sulfotransferase-3 protein sequence.
DE	
XX	Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
XX	selectin binding interaction; inflammation; lymphocyte homing; human;
KW	secondary lymph organ.
KM	
XX	
OS	Homo sapiens.
XX	
PN	W09949018-A1.
XX	
PD	30-SEP-1999.
XX	
PF	26-FEB-1999; 99WO-US04316.
XX	
PR	20-MAR-1998; 98US-0045284.
XX	PR 12-NOV-1998; 98US-0190911.
PA	(SYNTC) UNIV CALIFORNIA.
XX	(REGC) SYNTAX USA INC.
XX	
PI	Bistrup A, Rosen SD, Tangemann K, Hemmerlich S;
XX	WPI; 1999-580442/49.
DR	N-PSDB; AA20792.

XX Human and murine glycosyl sulfotransferase 3 and related
 PT polynucleotides
 XX
 PS Claim 2; Fig 1; 59pp; English.
 CC This sequence is the human glycosyl sulfotransferase-3 (GST-3) of
 CC the invention. The nucleic acid sequences, probes and primers derived
 CC from these, proteins and antibodies are useful in detecting homologues.
 CC The sequences, antibodies and methods are useful in the diagnosis and
 CC treatment of diseases associated with selectin binding interactions,
 CC including conditions associated with or resulting from the homing of
 CC leukocytes to sites of inflammation and the normal homing of lymphocytes
 CC to secondary lymph organs.
 CC
 XX
 SQ Sequence 386 AA;
 Query Match 100.0%; Score 2038; DB 20; Length 386;
 Best Local Similarity 100.0%; Pred. No. 1.3e-205;
 Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLTPKMKLLFLVSQMAITLFFHMYSHNSSLKMAQPERMHVILSSMRGSSPVG 60
 Db 1 MLTPKMKLLFLVSQMAITLFFHMYSHNSSLKMAQPERMHVILSSMRGSSPVG 60
 QY 61 LFGQHPDVFYLMRPANHWMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQ 120
 Db 61 LFGQHPDVFYLMRPANHWMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQ 120
 QY 121 SSLFQWNSRALCSAPACDIIPODEIIPRAHCRILCSQPFEEVERACRSYSHVLEVR 180
 Db 121 SSLFQWNSRALCSAPACDIIPODEIIPRAHCRILCSQPFEEVERACRSYSHVLEVR 180
 QY 181 FFMLOSLYPLKDPSPSLNHLVLRDPRAVFRSREKTKGMDLSRIYMGQHQKLKED 240
 Db 181 FFMLOSLYPLKDPSPSLNHLVLRDPRAVFRSREKTKGMDLSRIYMGQHQKLKED 240
 QY 241 QPYVMQVTCOSQLEIYKTIQSLPKAOERYLLVREEDLARAVAOISRYEFGLEFLP 300
 Db 241 QPYVMQVTCOSQLEIYKTIQSLPKAOERYLLVREEDLARAVAOISRYEFGLEFLP 300
 QY 301 HLOTWVHNITRGKMGDHAFTNARDALNVSOAMRWSLPYEKYSRLQKACGDMNLGTR 360
 Db 301 HLOTWVHNITRGKMGDHAFTNARDALNVSOAMRWSLPYEKYSRLQKACGDMNLGTR 360
 QY 361 HVRSEQQRNLIDLSTWTPPEQIH 386
 Db 361 HVRSEQQRNLIDLSTWTPPEQIH 386
 RESULT 2
 AAM93309 standard; Protein; 386 AA.
 XX
 AC AAM93309;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide, SEQ ID NO: 2817.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping.
 XX
 OS Homo sapiens.
 XX
 PN EPI130094-A2.
 PD 05-SEP-2001.
 PF 07-JUL-2000; 2000EP-0114089.
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.

XX (HELI-) HELIX RES INST.
 PA Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y.
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX WPI; 2001-524255/58.
 DR N-PSDB; AAK94229.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 PS Claim 8; SEQ ID NO 2817; 1380pp + sequence listing; English.
 CC
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a polypeptide
 CC encoded by a full length human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 CC
 XX
 SQ Sequence 386 AA;
 Query Match 99.9%; Score 2035; DB 22; Length 386;
 Best Local Similarity 99.7%; Pred. No. 2.7e-205;
 Matches 385; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLTPKMKLLFLVSQMAITLFFHMYSHNSSLKMAQPERMHVILSSMRGSSPVG 60
 Db 1 MLTPKMKLLFLVSQMAITLFFHMYSHNSSLKMAQPERMHVILSSMRGSSPVG 60
 QY 61 LFGQHPDVFYLMRPANHWMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQ 120
 Db 61 LFGQHPDVFYLMRPANHWMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQ 120
 QY 121 SSLFQWNSRALCSAPACDIIPODEIIPRAHCRILCSQPFEEVERACRSYSHVLEVR 180
 Db 121 SSLFQWNSRALCSAPACDIIPODEIIPRAHCRILCSQPFEEVERACRSYSHVLEVR 180
 QY 181 FFMLOSLYPLKDPSPSLNHLVLRDPRAVFRSREKTKGMDLSRIYMGQHQKLKED 240
 Db 181 FFMLOSLYPLKDPSPSLNHLVLRDPRAVFRSREKTKGMDLSRIYMGQHQKLKED 240
 QY 241 QPYVMQVTCOSQLEIYKTIQSLPKAOERYLLVREEDLARAVAOISRYEFGLEFLP 300
 Db 241 QPYVMQVTCOSQLEIYKTIQSLPKAOERYLLVREEDLARAVAOISRYEFGLEFLP 300
 QY 301 HLOTWVHNITRGKMGDHAFTNARDALNVSOAMRWSLPYEKYSRLQKACGDMNLGTR 360
 Db 301 HLOTWVHNITRGKMGDHAFTNARDALNVSOAMRWSLPYEKYSRLQKACGDMNLGTR 360
 QY 361 HVRSEQQRNLIDLSTWTPPEQIH 386
 Db 361 HVRSEQQRNLIDLSTWTPPEQIH 386
 RESULT 3
 AAU11274 standard; Protein; 380 AA.
 XX
 AC AAU11274;
 XX
 DT 12-MAR-2002 (first entry)
 XX
 DE Human L-selectin sulfotransferase-2 (LSS-2) protein.
 XX
 KW Human; beta1,3GNT; beta1,3-N-acetylglucosaminyltransferase; MECA-79;

KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
 KW ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;
 KW allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;
 KW delayed-type hypersensitivity reaction; hyperplastic thymus; antileuc;
 KW antiinflammatory; antipsoriatic; antidiabetic; dermatological;
 KW antiallergic.
 XX Homo sapiens.
 XX OS
 XX MO200185177-A1.
 XX
 XX PD 15-NOV-2001.
 XX
 XX PF 10-MAY-2001; 2001WO-US15452.
 XX
 XX PR 11-MAY-2000; 2000US-0569320.
 XX
 XX (BURN-) BURNHAM INST.
 XX
 XX PI Fukuda M, Yeh J, Hiraoka N;
 XX
 XX DR WPI: 2002-075226/10.
 XX DR N-PSDB; AAS16947.
 XX
 XX PT New enzyme, useful for modifying acceptor molecule, comprises an
 XX isolated L-selectin sulfotransferase-2 that directs expression of
 XX L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
 XX intestinal G10MC 6-sulfotransferase
 XX
 XX PS Claim 21; Fig 4; 98pp; English.
 XX
 XX The present invention provides a method of modifying an acceptor molecule
 XX by contacting the acceptor with an isolated
 XX betafal,3-N-acetylglucosaminyltransferase (betafal,3GNT) or an active
 XX fragment, where betafal,3GNT directs expression of a MECA-79 antigen. The
 XX invention also provides a method of treating or preventing an
 XX L-selectin-mediated condition by reducing the expression or activity of a
 XX betafal,3GNT that directs expression of a MECA-79 antigen. This can be done
 XX by administering to the subject an oligosaccharide L-selectin antagonist
 XX that inhibits binding of L-selectin to a MECA-79 antigen, for example by
 XX administering antibody material that specifically binds betafal,3GNT,
 XX and/or a betafal,3GNT antisense nucleic acid molecule. L-selectin
 XX sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.
 XX Alternatively, the expression or activity of LST-2 or its active
 XX fragment can be reduced in combination with reducing the expression or
 XX activity of betafal,3GNT. The method is useful for treating L-selectin
 XX mediated conditions such as Crohn's disease and ulcerative colitis,
 XX inflammatory disorders of the skin such as allergic contact dermatitis,
 XX psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
 XX hypersensitivity reactions, diabetes and hyperplastic thymus. This
 XX sequence represents human LST-2.
 XX
 SQ Sequence 380 AA;
 Query Match 98.5%; Score 2008; DB 23; Length 380;
 Best Local Similarity 100.0%; Pred. No. 1.8e-202;
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 QVCSOLEIKTIOSLPKALOERYLLVREYEDLAPAPVAQTSRMTEFVGLFPLPHIQTV 306
 DB 241 QVCSOLEIKTIOSLPKALOERYLLVREYEDLAPAPVAQTSRMTEFVGLFPLPHIQTV 300
 QY 307 HNITRGKMGDAHFTNARDALNVSAWRMSLPYKVSRLQKACGDANMLGRRHRSQ 366
 DB 301 HNITRGKMGDAHFTNARDALNVSAWRMSLPYKVSRLQKACGDANMLGRRHRSQ 360
 QY 367 EQRNLLDLSTWTPPEQIH 386
 DB 361 EQRNLLDLSTWTPPEQIH 380
 RESULT 4
 ID AAY79219 standard; Protein, 386 AA.
 AC AAY79219;
 XX
 XX 19-JUN-2000 (first entry)
 XX
 DE Human transferase TRNSFS-11.
 XX
 XX Transferase; TRNSFS-11; human; antitumor; cell proliferation;
 KW gastrointestinal disorder; developmental disorder;
 KW genetic disorder; neurological disorder; reproductive disorder;
 KW smooth muscle disorder; immunological disorder; inflammation;
 KW diagnosis; therapy; N-acetylglucosamine 6-O-sulfotransferase.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FT Modified-site 121
 FT Modified-site 107
 FT Modified-site 107
 FT Modified-site 217
 FT Modified-site 252
 FT Modified-site 364
 FT Modified-site 380
 FT Modified-site 35
 FT Modified-site 50
 FT Modified-site 81
 FT Modified-site 287
 FT Modified-site 243
 FT Modified-site 30
 FT Modified-site 308
 FT Modified-site 329
 FT Domain 7..23
 FT /note- "transmembrane domain"
 PN WO200014251-A2.
 XX
 XX 16-MAR-2000.
 XX PD 09-SEP-1999;
 XX PE 10-SEP-1998;
 XX PR 04-NOV-1998;
 PR 11-MAY-1999;
 98US-0150657.
 98US-0186779.
 99US-0133642.
 99WO-US20989.

XX (INCY-) INCYTE PHARM INC.
 XX Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;
 PI Hillman JL, Azimzal Y;
 XX MPI: 2000-256996/22.
 DR N-PSDB: AAZ94211.
 XX Human transferase proteins useful for preventing, diagnosing and
 PT treating cancers and developmental, gastrointestinal, genetic,
 PT immunological, neurological, reproductive and smooth muscle disorders -
 XX
 PS Claim 1; Page 90-91; 113pp; English.
 XX
 CC The present sequence is that of human transferase TRNSFS-11, 1 of
 CC 15 claimed novel human transferase proteins of the invention (see
 CC AAY79209-23). The sequence was deduced from a cDNA clone (see
 CC AAZ94211) isolated from a galbladder library. It shows homology to
 CC mouse N-acetylglucosamine 6-O-sulfotransferase. TRNSFS-11 is
 CC expressed in dermatologic and gastrointestinal tissues,
 CC especially those associated with inflammation and cell
 CC proliferation. The new human transferases and polynucleotides can
 CC be used in the diagnosis, prevention and treatment of cancer,
 CC developmental disorders, gastrointestinal disorders, genetic
 CC disorders, immunological disorders, neurological disorders,
 CC reproductive disorders, and smooth muscle disorders. The
 CC polypeptides can also be used to raise antibodies, and to screen
 CC for agonists and antagonists of transferase activity.
 CC
 SQ Sequence 386 AA;
 XX
 Query Match 95.0%; Score 1936; DB 21; Length 386;
 Best Local Similarity 93.6%; Pred. No. 7, 2e-195;
 Matches 369; Conservative 3; Mismatches 14; Indels 0; Gaps 0;
 QY 1 MLTPKMKLLFLVSOMATLALFFHMYSHNLSLSKMAOERHMYLVLSMRGSSPVG 60
 DB 1 MLTPKMKLLFLVSOMATLALFFHMYSHNLSLSKMAOERHMYLVLSMRGSSPVG 60
 QY 61 LFGQHPDVFLMEPAMHVMTEFKOSTAMMLHMAVNDLIRAVFLCDMSVFDAYMEGPRRQ 120
 DB 61 LFGQHPDVFLMEPAMHVMTEFKOSTAMMLHMAVNDLIRAVFLCDMSVFDAYMEGPRRQ 120
 QY 121 SSLFQWENSRALCSAPACDIIPODEITIPRAHCRLLCSQPFVEVEKACRSYSHVYLKEVR 180
 DB 121 SSLFQWENSRALCSAPACDIIPODEITIPRAHCRLLCSQPFVEVEKACRSYSHVYLKEVR 180
 QY 181 FENLQSLVPLKDPSTLNLHIVLRDPRAVFRSRETKGDLMDSRITVMGQHEOKLKED 240
 DB 181 FENLQSLVPLKDPSTLNLHIVLRDPRAVFRSRETKGDLMDSRITVMGQHEOKLKED 240
 QY 241 QPYVMQVICOSELETKTQSLPKALQERYLLVREDELAARVPAOTSRYMEFEVLEFLP 300
 DB 241 QPYVMQVICOSELETKTQSLPKALQERYLLVREDELAARVPAOTSRYMEFEVLEFLP 300
 QY 301 HLOTWVHNTTRGKGMDHAFHTNARDALNVSQAMPRLPYEKYSRLQKACGDMNLGTYR 360
 DB 301 HLOTWVHNTTRGKGMDHAFHTNARDALNVSQAMPRLPYEKYSRLQKACGDMNLGTYR 360
 QY 361 HVRSQEQORNLIDLSTWTVPEQIH 386
 DB 361 HVRSQEQORNLIDLSTWTVPEQIH 386
 RESULT 5
 AAY39919 standard; Protein; 388 AA.
 ID AAY39919 standard; Protein; 388 AA.
 AC AAY39919;
 XX 08-DEC-1999 (first entry)

DE Mouse glycosyl sulfotransferase-3 protein sequence.
 XX Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
 XX selectin binding interaction; Inflammation; lymphocyte homing; mouse;
 KM secondary lymph organ.
 XX
 XX Mus sp.
 XX MO9949018-A1.
 XX 30-SEP-1999.
 XX 26-FEB-1999; 99MO-0504316.
 XX 20-MAR-1998; 98US-0045284.
 XX 12-NOV-1998; 98US-0190911.
 PA (REGC) UNIT CALIFORNIA.
 PA (SYNT) SYNTAX USA INC.
 PI Bistrup A, Rosen SD, Tangemann K, Hemmerlich S;
 DR MPI: 1999-580442/49.
 DR N-PSDB: AAZ20793.
 XX
 PT Human and murine glycosyl sulfotransferase 3 and related
 PT polynucleotides
 PS Claim 2; Fig 4; 59pp; English.
 XX
 CC This sequence is the mouse glycosyl sulfotransferase-3 (GST-3) of
 CC the invention. The nucleic acid sequences, probes and primers derived
 CC from these, proteins and antibodies are useful in detecting homologues.
 CC The sequences, antibodies and methods are useful in the diagnosis and
 CC treatment of diseases associated with selectin binding interactions,
 CC including conditions associated with or resulting from the homing of
 CC leukocytes to sites of inflammation and the normal homing of lymphocytes
 CC to secondary lymph organs.
 CC
 SQ Sequence 388 AA;
 XX
 Query Match 74.3%; Score 1513.5; DB 20; Length 388;
 Best Local Similarity 72.8%; Pred. No. 2, 2e-150;
 Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;
 QY 1 MLTPKMKLLFLVSOMATLALFFHMYSHNLSLSKMAOERHMYLVLSMRGSSPVG 59
 DB 1 MLTPKMKLLFLVSOMATLALFFHMYSHNLSLSKMAOERHMYLVLSMRGSSPVG 58
 QY 60 QLEGOHPDVFLMEPAMHVMTEFKOSTAMMLHMAVNDLIRAVFLCDMSVFDAYMEGPRR 119
 DB 60 QLEGOHPDVFLMEPAMHVMTEFKOSTAMMLHMAVNDLIRAVFLCDMSVFDAYMEGPRR 118
 QY 59 QLEGOHPDVFLMEPAMHVMTEFKOSTAMMLHMAVNDLIRAVFLCDMSVFDAYMEGPRR 118
 DB 59 QLEGOHPDVFLMEPAMHVMTEFKOSTAMMLHMAVNDLIRAVFLCDMSVFDAYMEGPRR 117
 QY 120 QSSLFQWENSRALCSAPACDIIPODEITIPRAHCRLLCSQPFVEVEKACRSYSHVYLKEVR 179
 DB 120 QSSLFQWENSRALCSAPACDIIPODEITIPRAHCRLLCSQPFVEVEKACRSYSHVYLKEVR 178
 QY 179 RFLSLQALVPLKDPSTLNLHIVLRDPRAVFRSRETKGDLMDSRITVMGQHEOKLKED 238
 DB 179 RFLSLQALVPLKDPSTLNLHIVLRDPRAVFRSRETKGDLMDSRITVMGQHEOKLKED 237
 QY 240 DQPYVMQVICOSELETKTQSLPKALQERYLLVREDELAARVPAOTSRYMEFEVLEFLP 299
 DB 240 DQPYVMQVICOSELETKTQSLPKALQERYLLVREDELAARVPAOTSRYMEFEVLEFLP 298
 QY 300 PHLOTWVHNTTRGKGMDHAFHTNARDALNVSQAMPRLPYEKYSRLQKACGDMNLGTYR 359
 DB 300 PHLOTWVHNTTRGKGMDHAFHTNARDALNVSQAMPRLPYEKYSRLQKACGDMNLGTYR 358
 QY 360 HVRSQEQORNLIDLSTWTVPEQI 385
 DB 360 HVRSQEQORNLIDLSTWTVPEQI 384

CC The present invention describes human corneal
 CC N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyse
 CC sulfation of keratan sulfate (KS). Also described is a method for
 CC monitoring the effect of treatments for macular corneal dystrophy (MCD),
 CC and detecting susceptibility to MCD. (I) is located to chromosome 16q22,
 CC and has ophthalmological activity. (I) can be used to treat or prevent
 CC macular corneal dystrophy types I or II. (I) makes possible treatment
 CC of MCD without requiring keratoplasty or keratectomy. The present
 CC sequence represents mouse intestinal N-acetylglucosamine-6-
 CC sulfotransferase, which is given in comparison with (I) in the
 CC exemplification of the present invention.

SO Sequence 418 AA:

Query Match 50.4%; Score 1028; DB 23; Length 418;
 Best Local Similarity 56.6%; Pred. No. 3.2e-99;
 Matches 215; Conservative 43; Mismatches 112; Indels 10; Gaps 6;

QY 1 MLTPK--KMKLLFLYSQAMAILALFFHMTSHNLSLSMKAPRMVTVLSSRRSSSY 58
 DB 25 MRLPRSSVTWMLSLWVQGLVFE--LVSROVPS--SPAGLGRVHVLVLSMRSSSPV 80
 QY 59 GOLFGQHPDVFYLMPEAMHVMWTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGPR 118
 DB 81 GOLFGQHPDVFYLMPEAMHVMWTLSSGSAFALHMAVRDLIRSVFLCDMDVFDAYL-PMRR 139
 QY 119 ROSSLFQWENSRLCSAPACDIIPODEIIPRAHCRLLCSQOPEVEVEKACRSYSHVYLKE 178
 DB 140 NISDLFQWAVSRALCSPPVCEAFARCNISSEVECKFLCATRPGLAOCSSSYSHVYLKE 199
 QY 179 VRFENQSLYPLIKDLSLMLHTVLRDPRAVFRSREKRGDMLDSRYMGHEQKLK 238
 DB 200 VRFENQSLYPLISDPALMLRIVHLRDPRAVLRSEQFAKALARNGVLTGTNGTW- E 258
 QY 239 EDQPYVMOVICSOSELEIKTQSLPKALQERYLTVREEDLAPARVQTSRMFEVGLER 298
 DB 259 ADRPLKAVNEVCSHRVRIEALHKKPPFLODRRLVRYEDLADPLTVIRELTAFTGLGL 318
 QY 299 LPHLOTVHNITRGKMG--DHAFTNARDALNVSQAMRWSLPEKVSRLQKACGDAMNL 356
 DB 319 TPOLOTVHNITRGSGPARREAFKTSRDALSVQAMRWHTLPFAKIRVQELCGALQL 378
 QY 357 LGYRHVRSQEOQRNLDDL 376
 DB 379 LGYRHVSELEQRDLSDDL 398

RESULT 8
 AAY72640
 ID AAY72640 standard; Protein; 395 AA.

AC AAY72640;
 DT 02-MAY-2001 (first entry)

DE Human glycosyl sulfotransferase-4beta (GST-4beta).

XX Human; glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive;
 XX therapy; selectin binding inhibitor; gene therapy; inflammation;
 XX systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 XX glomerulonephritis; myasthenia gravis; Sjogren's syndrome; dermatitis;
 XX Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 XX demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 XX myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 XX asthma; hypersensitivity; rheumatic fever; tissue rejection;
 XX chromosome 16q23.1.

XX Homo sapiens.
 XX OS
 XX PN WO200106015-A1.
 XX PD 25-JAN-2001.

XX 19-JUL-2000; 2000MO-0519741.
 XX 20-JUL-1999; 99US-0144694.
 XX 13-JUL-2000; 2000US-0593828.
 XX (REGC) UNIV CALIFORNIA.
 XX Rosen SD, Lee JK, Hemmerich S;
 XX WPI: 2001-138471/14.
 XX DR N-FSDB; AAD02697, AAD02700.
 XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 XX diagnostic and therapeutic agent screening applications
 XX claim 3; Fig 4B; 128pp; English.

CC The present sequence is human glycosyl sulfotransferase-4beta (GST-4
 CC beta). GST-4 gene is found on chromosome 16q23.1.
 CC GST is a type 2 membrane protein useful for inhibiting a binding event
 CC between a selectin and a selectin ligand, which comprises contacting the
 CC selectin with a non-sulphated selectin ligand. GST and a small molecular
 CC agent that inhibits the sulphation activity of GST. GST is also useful
 CC in inhibiting a selectin mediated binding event. GST is useful in gene
 CC therapy to treat disorders such as acute or chronic inflammation,
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adenitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation.

SO Sequence 395 AA:

Query Match 50.0%; Score 1019; DB 22; Length 395;
 Best Local Similarity 54.5%; Pred. No. 2.6e-98;
 Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

QY 2 LTPKMKLLFLYSQAMAILALFFHMTSHNLSLSMKAPRMVTVLSSRRSSSYVGL 61
 DB 14 LLLAQFLLFLVSRP-----GPSSPAGGEARVHVLVLSRRSSSPVGL 59
 QY 62 FGOHPDVFYLMPEAMHVMWTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGPR 121
 DB 60 FNOHPDVFYLMPEAMHVMWTLSSGSAATLHMAVRDLIRSVFLCDMDVFDAYL-PMRRNS 118
 QY 122 SLFQWENSRLCSAPACDIIPODEIIPRAHCRLLCSQOPEVEVEKACRSYSHVYLKEVR 181
 DB 119 DLFQWAVSRALCSPPVCEAFARCNISSEVECKFLCATRPGLAOCSSSYSHVYLKEVR 178
 QY 182 FNIQSLYPLIKDLSLMLHTVLRDPRAVFRSREKRGDMLDSRYMGHEQKLK 241
 DB 179 FNIQSLYPLISDPALMLRIVHLRDPRAVLRSEQFAKALARNGVLTGTNGTW- EADP 237
 QY 242 PYVMOVICSOSELEIK--TIOSLPKALQERYLTVREEDLAPARVQTSRMFEVGLER 299
 DB 238 GLRVAVNEVCSHRVRIEALHKKP-PPFLGRRLVRYEDLADPLTVIRELTAFTGLSL 296
 QY 300 PHLOTVHNITRGKMG--DHAFTNARDALNVSQAMRWSLPEKVSRLQKACGDAMNL 357
 DB 297 PQLQVAVNHITRGSGPARREAFKTSRRNALNVSQAMRWHTLPFAKIRVQELCGALQL 356
 QY 358 GYRHVRSQEOQRNLDDL-----STVTPVQIH 386
 DB 357 GYRPVYSEDEQRNLDDLVLPRGLNGFTWASSTASH 392

RESULT 9
 ABB81554

```

ID  ABB81554 standard; Protein: 395 AA.
XX
AC  ABB81554:
XX
DT  05-SEP-2002 (first entry)
XX
DE  Human corneal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:2.
XX
KW  Human: N-acetylglucosamine-6-sulfotransferase; enzyme: GICNAC6ST;
KM  corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KM  ophthalmological; chromosome 16q22.
XX
OS  Homo sapiens.
XX
PN  US2002061562-A1.
XX
PD  23-MAY-2002.
XX
PF  09-AUG-2001; 2001US-0927602.
XX
PR  11-AUG-2000; 2000US-325773P.
XX
PA  (FUKU/) FUKUDA M N.
PA  (AKAM/) AKAMA T O.
XX
PI  Fukuda MN, Akama TO;
XX
DR  WPI: 2002-507643/54.
DR  N-PSDB: ABB81554.
XX
PT  New nucleic acid encoding corneal
PT  N-acetylglucosamine-6-sulfotransferase, useful for treatment,
PT  monitoring and diagnosis of macular corneal dystrophy
XX
PS  Claim 13; Fig 1A-D; 69pp; English.
XX
CC  The present sequence represents human corneal
CC  N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyse
CC  sulfation of keratan sulfate (KS). Also described is a method for
CC  monitoring the effect of treatments for macular corneal dystrophy (MCD),
CC  and detecting susceptibility to MCD. (I) is located to chromosome 16q22,
CC  and has ophthalmological activity. (I) can be used to treat or prevent
CC  macular corneal dystrophy types I or II. (I) makes possible treatment
CC  of MCD without requiring keratoplasty or keratectomy.
XX
SQ  Sequence 395 AA:
Query Match 50.0%; Score 1019; DB 23; Length 395;
Best Local Similarity 54.5%; Pred. No. 2,66-98;
Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;
QY  2 LLPKMKLLLEFLVSGMAIILFFHMYSHNISLSMKAPERHMYLVLSMSSGSSFFVQL 61
DB  14 LLAAQTLELFLVSRP-----GSSPAGEAVHVAHVHVLSSMSSGSSFFVQL 59
QY  62 FGHDPDFVYLMPEAMHWMTFKOSTAWMLHMAVBDLIRAVELCDMSYFDAYMEGPRROS 121
DB  60 FNGHPDFVYLMPEAMHWMTFKOSTAWMLHMAVBDLIRAVELCDMSYFDAYMEGPRROS 118
QY  122 SLFWENSRALCSAPACDIIPODEIIPRAHCRLLCSQPFVEVERACRSYSHVVLKEVRF 181
DB  119 DLFQWAVSRALCSAPACDIIPODEIIPRAHCRLLCSQPFVEVERACRSYSHVVLKEVRF 178
QY  182 FNLQSLVPLKDPSSLNHYVLVDRPRAVRSRRTGDMIDRSRYVGHQEQILKKEDQ 241
DB  179 FNLQSLVPLKDPSSLNHYVLVDRPRAVRSRRTGDMIDRSRYVGHQEQILKKEDQ 237
QY  242 PYVWQVYICQSOLEIYK--TIQSLPKALQERYLVREYEDLARAVAQSRMYEVEGLEFL 299
DB  238 GLRVYREVCRSHVAIAEAATIKP--PPLRGYRVLVREYEDLARAVAQSRMYEVEGLEFL 296
QY  300 PHLOTWVHNITRGKMG--DHAFTNADALNVQAWMSLPEYKVSRLQACDAMNLT 357
DB  357 PHLOTWVHNITRGKMG--DHAFTNADALNVQAWMSLPEYKVSRLQACDAMNLT 357

```

```

DB  297 PQLFAMHNHTHSGPGARREAFKTSRRNALNVQAWRHLPFAKIRRHQELCAGALQL 356
QY  358 GYRVRSQEQEORNLIDL-----STWYVPEQIH 386
DB  357 GYRVYSDEQRNALDLVLPRLGLNGFTWASSTASH 392
RESULT 10
ID  AAE15438
ID  AAE15438 standard; Protein: 395 AA.
XX
AC  AAE15438;
XX
DT  12-MAR-2002 (first entry)
XX
DE  Human drug metabolizing enzyme (DME)-5.
XX
KW  Human; drug metabolizing enzyme; gene therapy; autoimmune disorder;
KW  inflammatory disorder; acquired immune deficiency syndrome; infection;
KW  AIDS; adult respiratory distress syndrome; cell proliferative disorder;
KW  allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;
KW  cancer; endocrine disorder; hypothyroidism disorder; pituitary disorder;
KW  gastrointestinal disorder; metabolic disorder; developmental disorder;
KW  liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;
KW  goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
KW  DME-5.
XX
OS  Homo sapiens.
XX
XX
XX  Key Location/Qualifiers
XX  Peptide 1..32
XX  Protein /label=Signal_peptide
XX  Peptide /note="Human mature DME-5 protein"
XX  Protein /label=Signal_peptide
XX  Protein /note="Human mature DME-5 protein"
XX
XX  WO200179468-A2.
XX
XX  25-OCT-2001.
XX
XX  12-APR-2001; 2001WO-US11869.
XX
XX  13-APR-2000; 2000US-197590P.
XX  19-APR-2000; 2000US-198403P.
XX  28-APR-2000; 2000US-200185P.
XX  05-MAY-2000; 2000US-20234P.
XX  11-MAY-2000; 2000US-203509P.
XX
XX  (INCY-) INCYTE GENOMICS INC.
XX
XX  Policky JT, Hafalia A, Burford N, Ring HZ, Lal P, Tribouley CM;
XX  Yao MG, Yue H, Tang YT, Patterson C, Das D, Sanjwalala MS;
XX  Gandhi AR, Reddy R, Khan FA, Baughn MR, Ramkumar J, Griffin JA;
XX  Au-Young J;
XX
XX  WPI: 2002-066363/09.
XX  N-PSDB: AAD24670.
XX
XX  Novel isolated human drug metabolizing enzymes referred as DME 1-10,
XX  useful for diagnosing, treating, or preventing disorders associated
XX  with aberrant expression of DME such as allergy, anemia, asthma,
XX  infertility
XX
XX  Claim 1a; Page 131-132; 143pp; English.
XX
XX  The invention relates to human drug metabolizing enzymes referred as
XX  DME and nucleic acid molecules encoding such enzymes. Polynucleotides
XX  of the invention are useful for assessing toxicity of test compounds
XX  and in gene therapy. Sequences of the invention are useful in the
XX  diagnosis, prevention and treatment of autoimmune/inflammatory

```


CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation.

CC Note: The present sequence is also shown in sequence listing (page
CC no: 56) but lacks four nucleotides at its 3' end.

XX Sequence 390 AA;

Query Match 49.5%; Score 1008; DB 22; Length 390;
Best Local Similarity 52.3%; Pred. No. 3.7e-97;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

QY 1 MLPRKKMLLFLVSOAIIALFFHMYSHNISLSMKAPRPMHVLVLSWRSGSSFFVQ 60
DB 14 LLLAQTTCLLFFISRP-----GPSSPAGEDRVHVLVLSWRSGSSFFVQ 59
QY 61 LFGQHPDVFYLMERPAHMYMTFKOSTAMLMNAVRLIRAVFLCDMSYFDAYMEGPRQ 120
DB 60 LFSQHPDVFYLMERPAHMYMTFKOSTAMLMNAVRLIRAVFLCDMSYFDAYM-POSRL 118
QY 121 SSLFQWENSRLCSAPACDIIPODEIIPRAHCRLLCSQOPEVEVEKACRSYSHVLYKEVR 180
DB 119 SAFNMTATSRALCSPPACSAFPRGTISKQDVCKTLCTQPSLAREACRSYSHVLYKEVR 178
QY 181 FFNIQSLYPLKDSNLNLIYHLYVDRPRAVRSERKRGDLMDSRYMGQEQKLKED 240
DB 179 FFNIQSLYPLKDSNLNLIYHLYVDRPRAVRSERKRGDLMDSRYMGQEQKLKED 237
QY 241 QPYVMQVICSQLEIK--TIQSLPKALOERYLLVREDAARAPVQTSMYEFGLEF 298
DB 238 PHLRLIREVCRSHVRIEAAATLKP--PPLRGRIYLVREDAARAPVQTSMYEFGLEF 296
QY 299 LPHLOTWVHNITRGKMGD--HAFHTNARDALNVSQAWMSLPYEKYSRLQKAGDAMNL 356
DB 297 TPQLEAWIHNTHTSGICKPIEAFHTSSRNARNVSQAWRHALLPFTKILRVQEVGAGALQL 356
QY 357 LGYRHVSEQQRNLLDLL-----STWTVPE 383
DB 357 LGYRHVSEQQRNLLDLL-----STWTVPE 390

RESULT 15

ABB81556
ID ABB81556 standard; Protein; 390 AA.

XX ABB81556;

XX 05-SEP-2002 (first entry)

DE Human intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:4.

KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GICNAC6ST;

KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;

XX opthlalmological.

XX Homo sapiens.

XX US2002061562-A1.

XX 23-MAY-2002.

XX 09-AUG-2001; 2001US-0927602.

XX 11-AUG-2000; 2000US-325773P.

XX (FUKU/) FUKUDA M N.

XX (AKAM/) AKAMA T O.

XX Fukuuda MN, Akama TO;

XX WPI; 2002-507643/54.

XX New nucleic acid encoding corneal

XX N-acetylglucosamine-6-sulfotransferase, useful for treatment.

PT monitoring and diagnosis of macular corneal dystrophy -
XX Example 5; Fig 2A-B; 69pp; English:

CC The present invention describes human corneal
CC N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyze
CC sulfation of keratan sulfate (KS). Also described is a method for
CC monitoring the effect of treatments for macular corneal dystrophy (MCD),
CC and detecting susceptibility to MCD. (I) is located to chromosome 16q22,
CC and has opthlalmological activity. (I) can be used to treat or prevent
CC macular corneal dystrophy types I or II. (I) makes possible treatment
CC of MCD without requiring keratoplasty or keratectomy. The present
CC sequence represents human intestinal N-acetylglucosamine-6-
CC sulfotransferase, which is given in comparison with (I) in the
CC exemplification of the present invention.

XX Sequence 390 AA;

Query Match 49.5%; Score 1008; DB 23; Length 390;
Best Local Similarity 52.3%; Pred. No. 3.7e-97;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

QY 1 MLPRKKMLLFLVSOAIIALFFHMYSHNISLSMKAPRPMHVLVLSWRSGSSFFVQ 60
DB 14 LLLAQTTCLLFFISRP-----GPSSPAGEDRVHVLVLSWRSGSSFFVQ 59
QY 61 LFGQHPDVFYLMERPAHMYMTFKOSTAMLMNAVRLIRAVFLCDMSYFDAYMEGPRQ 120
DB 60 LFSQHPDVFYLMERPAHMYMTFKOSTAMLMNAVRLIRAVFLCDMSYFDAYM-POSRL 118
QY 121 SSLFQWENSRLCSAPACDIIPODEIIPRAHCRLLCSQOPEVEVEKACRSYSHVLYKEVR 180
DB 119 SAFNMTATSRALCSPPACSAFPRGTISKQDVCKTLCTQPSLAREACRSYSHVLYKEVR 178
QY 181 FFNIQSLYPLKDSNLNLIYHLYVDRPRAVRSERKRGDLMDSRYMGQEQKLKED 240
DB 179 FFNIQSLYPLKDSNLNLIYHLYVDRPRAVRSERKRGDLMDSRYMGQEQKLKED 237
QY 241 QPYVMQVICSQLEIK--TIQSLPKALOERYLLVREDAARAPVQTSMYEFGLEF 298
DB 238 PHLRLIREVCRSHVRIEAAATLKP--PPLRGRIYLVREDAARAPVQTSMYEFGLEF 296
QY 299 LPHLOTWVHNITRGKMGD--HAFHTNARDALNVSQAWMSLPYEKYSRLQKAGDAMNL 356
DB 297 TPQLEAWIHNTHTSGICKPIEAFHTSSRNARNVSQAWRHALLPFTKILRVQEVGAGALQL 356
QY 357 LGYRHVSEQQRNLLDLL-----STWTVPE 383
DB 357 LGYRHVSEQQRNLLDLL-----STWTVPE 390

Search completed: August 9, 2003, 16:42:39
Job time : 41.3979 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:40:34 ; Search time 14.9612 Seconds
(without alignments)
2481.152 Million cell updates/sec

Title: US-09-645-078-2

Perfect score: 2038
Sequence: 1 MLPRKMKLLFLVSOMAIL.....EQRNLLDLSTWTFEQIH 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	623	30.6	484	2 JC7350	N-acetylglucosamin
2	606.5	29.8	486	2 JC7351	N-acetylglucosamin
3	588.5	28.9	484	2 JE0261	N-acetylglucosamin
4	500.5	24.6	458	2 A57397	chondroitin 6-sulf
5	168	8.2	307	2 E95934	probable enzyme, C
6	116	5.7	388	2 G70729	hypothetical prote
7	95	4.7	3588	2 I40485	surfactin syntheta
8	92.5	4.5	2354	2 T13288	met-41 protein - f
9	92	4.5	359	2 T16350	hypothetical prote
10	91.5	4.5	338	2 JE0196	hydroxysteroid sul
11	91	4.5	659	2 S11737	resistance protein
12	91	4.5	1847	2 E64477	replication factor
13	89.5	4.4	346	2 E86319	probable flavonol
14	89.5	4.4	504	2 T07120	probable cyclochrom
15	89.5	4.4	4273	2 C69679	polyketide synthas
16	89	4.4	524	1 A36205	mitochondrial proc
17	89	4.4	1140	2 G71335	probable transcrip
18	89	4.4	1968	1 PN0093	genome polypotein
19	88	4.3	359	2 T47187	hypothetical prote
20	88	4.3	466	2 E84132	aminopeptidase BH3
21	87.5	4.3	380	2 T42755	tyrosylprotein sul
22	87.5	4.3	4063	2 T42993	probable spectrin
23	87.5	4.3	4101	2 T23630	hypothetical prote
24	87	4.3	2500	2 G88493	protein E57B9.2 [1
25	86.5	4.2	748	1 S08680	methylnalonyl-CoA
26	86.5	4.2	831	2 S26675	DNA-directed DNA p
27	86.5	4.2	2352	2 T06077	splicing factor PR
28	86	4.2	320	2 A40216	flavonol 4'-sulfot
29	86	4.2	823	2 S48986	probable protein x

30	86	4.2	869	2 A96558	probable protein k
31	85.5	4.2	344	2 H83928	hypothetical prote
32	85.5	4.2	764	1 S14113	1-phosphatidylinos
33	85	4.2	632	2 T18692	hypothetical prote
34	85	4.2	672	2 T34899	probable DNA recom
35	85	4.2	1237	2 T37529	hypothetical prote
36	85	4.2	2688	2 I49477	alpha-A-crystallin
37	84.5	4.1	217	2 J01358	C1 protein - Mica
38	84.5	4.1	217	2 C86863	theonine-tRNA 119
39	84.5	4.1	1031	2 T43458	hypothetical prote
40	84.5	4.1	2493	2 S45734	probable membrane
41	84	4.1	390	2 AB2337	group 2 sigma 70-t
42	84	4.1	1189	2 T42726	guanine nucleotide
43	84	4.1	1217	2 T00270	hypothetical prote
44	84	4.1	1583	2 T14176	probable phosphati
45	83.5	4.1	518	2 E84536	hypothetical prote

ALIGNMENTS

RESULT 1

JC7350
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse
C:Species: Mus musculus (house mouse)
C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000
C:Accession: JC7350

R:Uchimura, K.; Fasaki, F.; Kadomatsu, K.; Matsukawa, T.; Yamakawa, T.; Kurosawa, A.; Bloem, B. Res. Commun. 274, 291-296, 2000

A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of A:Reference number: JC7350; MUID:20374462; PMID:10913333

A:Accession: JC7350

A:Residues: 1-484 <UCH>

A:Cross-references: DDBJ:AB040710

C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosyl sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 30.6% Score 623; DB 2; Length 484;

Best Local Similarity 34.2% Pred. No. 1.5e+45;

Matches 136; Conservative 65; Mismatches 151; Indels 46; Gaps 7;

OY	30	NISSLSMAKQPERHNVIVLSSMRSGSSPVGOLFQHPDVFYILMPAMHMTFKOSTAMM	89
DB	87	NLSAVGEVYDEKQHYHATWRTGSSSLGELFPHQHPVFYILPEPMHMLQALPGDAES	146
OY	90	LHMAVRDLIRAVFLCDMSVFDAYMEPGPRQ-----SSLEWENSRLCSAPACD	139
DB	147	LQALRDLRLSLFRCDFSVLRLVLAQPGDGRAPDSANLTYAMLEFRWRTNVICSPICP	206
OY	140	IIIPQDE-----IIPAHCRILCSQOPFEYVEKACSYSNVLKEVFFWLSLYPLADPS	195
DB	207	AAPRARADVGLVEKACESTCPVSLRLAEKCRYPVYVLIKDVRLDLGLVPLRLPG	266
OY	196	LNLHIVHLVRPRAVFRSRETKGDLMDL-----RIWV-----GOHE	233
DB	267	LNLKVVQLFRPRAVHNSRLSKROGLLESIQVLETRQGRGHFRVLALMGVDARPGQA	326
OY	234	QKLKEDQPYV-----VMQVICSQLEIYKTQSLPKALQERYLLVREEDLARAVAQTSR	289
DB	327	RALPSAPRADPFLSLALEVICEMWLRDLFTRGAPAMLRRTYLRREYDLVWQAOQARR	386
OY	290	MYEEVGLPEFLPLOTWVHNITRGKMG-DIAFPHNARDALVSAWKRSLPEYEVSRLOK	348
DB	387	LIRFSGRLTALAALDAFAFMNTRGSAYGADRFHLSARARAVVHVRRLSQEQVQVET	446
OY	349	ACGDAMNLIGYRAVRSQEOENLL-----LDLSTWTFY	381
DB	447	ACAPAMRLAYPRSGDEDRKTVREGETPLETKANMAY	484

RESULT 2

NC7351 N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human

C:Species: Homo sapiens (man)
C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000
C:Accession: J07351

R:Uchimura, K.; Fasanian, A.; Kadomatsu, K.; Matsukawa, T.; Yamakawa, T.; Kurosawa, N.; Bloch, M. Biochem. Res. Commun. 274, 291-296, 2000

A>Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a cDNA from human liver
A:Accession: NC7351
A:Reference number: J07350; M01D:2037462; PMID:1091333

A:Molecule type: mRNA
A:Residues: 1-486 <UCH>

A:Cross-references: DDBJ:AB0140711
C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation linkage formation. This enzyme is involved in synthesis of L-selectin ligand in C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match Best Local Similarity 29.8%; Score 606.5; DB 2; Length 486;
Matches 136; Conservative 56; Mismatches 133; Indels 43; Gaps 8;

OY 41 ERHNVLVSSRSGSSFFVGLGQHPDVFEYLMEPAHYMTFKSTAMMLMAVRDLIRA 100
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 100 EKHIVHAATWRTSSSEFLGFELNQHDPVFYEPMHLMQLALPGDAESLGALRDMLRS 159
OY 101 VEFCDMSVEFADYMPG-PRRQ-----SLLQWENSRLCSAPACDIIPDE----I 146
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 160 LEKDFSVLRLYPGPAPPAARPTANTLTALFRMTRNKVCISPLCPGARAEVGL 219
OY 147 IPRAHCRLCSQGEFEVENKACRSYSHVLEKEVFNFNLSLYPLKPSSLNLIVHLVD 206
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 220 VEDPACERSCPVAIRALEACRKYPVVVIDVALLDGLGVLPRLDRGCLKKVQLFRD 279
OY 207 PRAVRSRERTGDLMIDS-----RIVM-----GGHEQLKREDPOYY 244
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 280 PRAVNSRLSKRGCLLRSLTQVLRTROGRDFHRVLLAHGVGARGCGQSRLPAAPRADF 339
OY 245 ---VMQVIOCSOLEIKTTIOSLPKALOERYLLRYBDLAAPAQGSRMTEFGLEGFLP 300
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 340 FLTGALVELICAMLRDLLFAFGADPAWLRRRYLTRLYEDELVNAPQAQLRRLRFSGLRALA 399
OY 301 HIAQTWNHTNRKGGMG-DHAFTNARDALANYSQAMRSLPEKYSRLOKAGDAMNLIIGY 359
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 400 ALDAFALNMRTGAAYGADRPHLSARDAREVHAMRELSHQYRVAAACAPARMRLLAY 459
OY 360 RHVRSEGE 367
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 460 --PRSGEE 465

RESULT 3

J0E0261 N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human
C:Species: Homo sapiens (man)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 01-Mar-2002
C:Accession: J0E0261

R:Uchimura, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Fan, Q.W.; Mitsunaka, T.

J:Biochem. 124, 670-678, 1998

A>Title: Human N-Acetylglucosamine-6-O-sulfotransferase Involved in the biosynthesis of A:Reference number: J0E0261; MUID:98391845; PMID:9722682

A:Accession: JE0261
A:Molecule type: mRNA
A:Residues: 1-484 <UCH>

A:Cross-references: DDBJ:AB014679
C:Comment: This protein catalyzes the transfer of sulfur of sulfate from 3'-phosphoadenosine C:Superfamily: chondroitin 6-sulfotransferase
C:Keywords: sulfotransferase

Query Match Best Local Similarity 28.9%; Score 588.5; DB 2; Length 484;
Matches 131; Conservative 56; Mismatches 137; Indels 29; Gaps 6;

Query Match 24.6% Score 500.5; DB 2; Length 458;
Best Local Similarity 33.2% Pred. No. 4.5e-33;
Matches 120; Conservative 75; Mismatches 133; Indels 33; Gaps 13;

Query 34 LSMKAQEPBMHVLTSSMRSSGSPVGOIFGQHPDVFYLMBPAMHV--WMTEFKOSTAMMLH 91
Db 104 LGTAPPEPRRHVLTMAATRTGSSVFGEFNOGNIFLYLEPLMHTEKTVTFEGOGANAVG 163
Query 92 MAV--RDLIRAVFLCDMSVFDAYMEPGRRO--SLFQWENSRALCSAPACDIIPQ-DEI 146
Db 164 SALVYRDVQLLCLDYLIESFISPADEHLTALPRRSSHSLCEPVC--TPSLKY 221
Query 147 IPRAHCR-LICSQOPFEVEKACRSYHYVLKEVRFNDSLPPLKDPSLNHYHVLVDRPVRSR 205
Db 222 FEKHCXNRRRCGPNTITLAEACRRKOHMLKTVRIROLEFLPDLADPRIDRIIDVLV 281
Query 206 DPRAPVRSRETKGDMIDSRIVGHOEOKLK-----KEQOPYVWQVICS-QLE 255
Db 282 DPRVAVLSR-----MV--AFSGKTESMKKMAABEDAPLOEDE-VQRLKGCESIRLS 330
Query 256 IYKTIQSIPKALQERYLLVRYEDLAPAPVAQTSMEYFVLEFLPHLQTVWHNITRGKM 315
Db 331 AELGLRQ-PRVLRRRYMLVRYEDVAPARPLKALEMYFAFAPIHPTPOVEEIRIANTQAP-Q 388
Query 316 GDHAFTHTNARDALNVQAMRWSLPEYKVSRLQKACGDAMNLLGRHYVRSQEQEQRNILLDL 375

Query 41 RRMVILVLTSSGSSPVGOLFQGPVFLMEBPAMHVMTDFKOSTAMMLHMAVRODIRA 100
Db 117 KRHHMYFTTMRSSSSFFGELFNQNPVEFLTEPVMYVQKTLPGDAVSLQAGARMLISA 176
Query 101 VFLCDMSVFDAYMEPGRRO---SLFQWENSRALCSAPACDIIPQDEIIPRAHCR-L 155
Db 177 LYRCDLSVFOQLYSPAGSGRNLTLTGIFGATANKVVCSSPLCPAY -KEVYGLVDYDRK 235
Query 156 -CSQOPFEVEKACRSYHYVLKEVRFNDSLPPLKDPSLNHYHVLVDRPVRSR 214
Db 236 KCPQORLAREEECRKRYTLVKGVRFDVAVALPRLRDLALDKLTHLVDRPVAASR 295
Query 215 ERTGDMLSRIYM-----GOHEOKLKE-----DQPY--VMQVICS 252
Db 296 IRRNHGLIRRESLQVYVRSDRPRAHMRPFLLEAAGHLLAKKKGCVGRADYHALLGMAEYICNS 355
Query 253 QLEIYKTIQSIPKALQERYLLVRYEDLAPAPVAQTSMEYFVLEFLPHLQTVWHNITRG 312
Db 356 MAKTLQALPPMDLQGHYLVVRVEDLVGEVYKTLRRVYDFVGLIVSPEMEOPALNNTSG 415
Query 313 KGMDDHAFHTNARDALNVQAMRWSLPEYKVSRLQKACGDAMNLLGRHYVRSQEQEQRNLL 372
Db 416 SGSSSKRPVVSARNATQANAMRTALTFOQIKOVEECYCPMAVLGIERVNSPEEVADLS 475
Query 373 LDLT 376
Db 476 KTLT 479

RESULT 4
A57397
chondroitin 6-sulfotransferase (EC 2.8.2.17) precursor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 20-Jun-2000
C:Accession: A57397
R:Fukuta, M.; Uchimura, K.; Nakashima, K.; Kato, M.; Kimata, K.; Shinomura, T.; Habuc
J. Biol. Chem. 270, 18575-18580, 1995
A:Title: Molecular cloning and expression of chick chondrocyte chondroitin 6-sulfotransferase
A:Reference number: A57397; PMID:95355490; PMID:7629189
A:Accession: A57397
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-458 <FRK>
A:Cross-references: GB:D49915; NID:g9711262; PIDN:BA008655.1; PID:g971263
C:Superfamily: chondroitin 6-sulfotransferase
C:Keywords: sulfotransferase

Db 389 DSNQIYSTOKNSSQEFKWRPSIFPKLAQVVDACAPAMRLFGYKLASSAQLTNRSL 448
 QY 376 L 376
 Db 449 L 449

RESULT 5

E95934

probable enzyme, C-terminal domain similar to sulfotransferase protein [Imported] - Sinc

C:Species: Sinorhizobium meliloti

C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: E95934

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo

A:Reference number: A95842; MUID:21396508; PMID:11481431

A:Accession: E95934

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-307 <KOR>

A:Cross-references: GB:AL1591985; PIDN:CAQ49141.1; PID:915140626; GSPDB:GN00167

A:Experimental source: strain 1021, megaplasmid pSymb

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,

heubalt, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: SMD21237

A:Genome: plasmid

Query Match 8.2%; Score 168; DB 2; Length 307;
 Best Local Similarity 22.7%; Pred. No. 9.2e-07;
 Matches 83; Conservative 50; Mismatches 131; Indels 102; Gaps 16;

QY 36 MKAQPERRHVLVLSWSSGSSFGQLFGQHPDVF--YLMPEAMHVMWTFKOSTAMLM 92
 Db 1 MFSQPVRL-AVIAQYGRSGSTILDLALGQHAAYVAGAGITSLTRHVMHNEYCA----- 53
 QY 93 AVRRLIRAVFLCDMSVVDAYMEPRPRSSSLFQWENSALCSAPACDIIIPDEIIIPRAHC 152
 Db 54 -----CGNAIRDCSFWSSVRR-----EWSDG-----QDPGLMEYC 84
 QY 153 RLCSQOPFEVEVEACR-----SYSHVLEKVEFFN-LQS----- 186
 b 85 AL---QKFEGLSMNTRLSLGSGMLGKQESLIILTKRFLFSAMQSCSGQVIVDSKLPGR 141
 QY 187 LYLPLKDPSLMLHVLVDRPRAVFRS-----RETKGIMLIDSRIVMGQHEQKKED 240
 Db 142 AMAVAQIPIGDMRIHVLVDRGVAMSLKGYERDAKSG-----LQKEI 185
 QY 241 QPVYVMOVQSOLEITKIOSLPKAL-QERYLLVRYEDLARAVAOQTSRMVEVEGFEFL 299
 Db 186 KPKSVFKRLALWMSM-VNLAVELYSRKLSGEVVMRREDESPVAVMQIGTFLDLD-L 243
 QY 300 PHLQTVNHNITRGKMG-----DHAFTNARDALNVSQAWMSLPEYKVSRLQKACGDA 353
 Db 244 SQVETSLEN---GEAMGQGHVAGVANNRLMNASIALNDEITWRTMRAQVSPQRLGGM 300
 QY 354 MNLIGY 359
 Db 301 LRRYGY 306

RESULT 6

G70729

hypothetical protein RV2267c - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: G70729

R:Coile, S.T.; Broesch, R.; Parthill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon

; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: G70729

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-388 <COD>

A:Cross-references: GB:Z77163; GB:AL123456; NID:93261610; PIDN:CAB00968.1; PID:e25507

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV2267c

Query Match 5.7%; Score 116; DB 2; Length 388;
 Best Local Similarity 19.3%; Pred. No. 0.036;
 Matches 70; Conservative 67; Mismatches 111; Indels 114; Gaps 20;

QY 45 VLVLSSWRSQSSFGVQLF--GQH--PDVFLMEPAHVMWTFKOSTAMLMAYRDLIR 99
 Db 83 IFIVGHMRTGTLHELLVVDDBRHTGPGYECIAP--HHFLT---TEM----- 125

QY 100 AVFLCDMSVFPAYMEPRRROSSLFQWENSALCSAPACDIIIPQ-DEII-----PRAHC 152
 Db 126 -----FAPYVE-----FLVSKHAMQMDLSLHNPDEDFWCMQGLPSPYL 167

QY 153 RLCSQOP-----FEVEKACRSYSHVLEKVEFFNLSLY-----PLKDPSLNHI 200
 Db 168 TIAPNRPPOYEYIDLEQVAPREL--IMKRTLFREFQOYVFRKRTVILKNPHSRI 225

QY 201 -----VHLVDRPRAVFRSRETKGIMLIDSRIVMGQHEQKKEDQPY--VM 246
 Db 226 KVLLEVPFQAKFIHIVDRPYVY-----PSYIHAKALYRIHGLOQPFEDGLD 273

QY 247 QVICOSQLEITKIOSLPKALQ-EKYLVRVEDLARAVAOQTSRMVEVEGFE-----FLPH 301
 Db 274 DKVASTYVDLRYKLDGRELVDPPREFYELRLELDLGDEGQRLRYQLRGAGDECCYLR 333

QY 302 LOTWVHNITRGKMGDHA-FHTNARDALNVSQAWMSLPEYKVSRLQKACGDAMNLIGY- 359
 Db 334 LRQY-----LADHADYKTN-----YQLTVQRAIVDEHMEIIDRYGYD 373

QY 360 RH 361
 Db 374 RH 375

RESULT 7

I40485

surfactin synthetase component I - Bacillus subtilis

N:Alternate names: competence protein srfAA; surfactin production protein srfAA; surf

C:Species: Bacillus subtilis

C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 03-Nov-2000

C:Accession: I40485; B69718; S35517; A37323; S46667; A43705; S34985

R:Cosmina, P.; Rodriguez, F.; de Ferra, F.; Grandi, G.; Perego, M.; Venema, G.; van S

Mol. Microbiol. 8, 821-831, 1993

A:Title: Sequence and analysis of the genetic locus responsible for surfactin synthe

A:Reference number: I40485; MUID:93360813; PMID:8355609

A:Accession: I40485

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-358 <RES>

A:Cross-references: EMBL:X70356; NID:9396480; PIDN:CAA49816.1; PID:9396481

A:Experimental source: strain W168 derivative of JH642

R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Ber

C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallazzi, A.; Galled
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koester, P.; Koningsstein, G.; Kirogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauesel
 Y, M.; Ogawa, K.; Ojivara, A.; Oudea, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelid
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
 akuchl, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wilpat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A:Reference number: A69580; MVID:98044033; PMID:9984377
 A:Accession: B69718
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-3588 <K0N>
 A:Cross-references: GB:299105; GB:AL009126; NID:g2632457; PIDN:CAB12142.1; PID:g2632634
 A:Experimental source: strain 168
 A:Title: Nucleotide sequence of 5' portion of *srfa* that contains the region required for
 Nucleic Acids Res. 21, 93-97, 1993
 A:Reference number: S35517; MVID:93181186; PMID:8441623
 A:Accession: S35517
 A:Status: nucleic acid sequence not shown; significant sequence differences
 A:Molecule type: DNA
 A:Cross-references: EMBL:D13262; NID:g216345; PID:g216346
 A:Experimental source: strain 168 tnpC2
 A:Note: protein sequence not complete, the nucleotide sequence was submitted to the EMBL
 R:Nakano, M.M.; Magnusson, R.; Myers, A.; Curry, J.; Grossman, A.D.; Zuber, P.
 J. Bacteriol. 173, 1770-1778, 1991
 A:Title: *srfa* is an operon required for surfactin production, competence development, an
 A:Reference number: A37323; MVID:91154134; PMID:1847909
 A:Accession: A37323
 A:Status: nucleic acid sequence not shown; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-46, 'EV', '49-145', 'I', '147-150', 'L', '152', 'AN', '155-280', 'T', '282-307', 'SF', '310-384'
 A:Cross-references: GB:M59939; NID:g14366
 R:Nakano, M.M.; Xia, L.; Zuber, P.
 J. Bacteriol. 173, 5487-5493, 1991
 A:Title: Transcription initiation region of the *srfa* operon, which is controlled by the
 A:Reference number: A43705; MVID:91358326; PMID:1715856
 A:Contents: annotation
 R:Fabret, C.; Quentin, Y.; Gutseppl, A.; Busuttil, J.; Halech, J.; Denizot, F.
 submitted to the EMBL Data Library, March 1993
 A:Reference number: S46967
 A:Accession: S46967
 A:Molecule type: DNA
 A:Residues: 3249-3271, 'A', '3273-3316', 'R', '3318-3451', 'Y', '3453-3483', 'DE', '3486-3487', 'DAGU', '34
 A:Cross-references: EMBL:X72672; NID:g516358; PIDN:CAA51222.1; PID:g516359
 A:Experimental source: strain 168 tnpC2
 C:Comment: This enzyme is one of several in the multienzyme complex that synthesizes the
 ne.
 C:Genetics:
 A:Gene: *srfa1*
 A:Note: *srfa1* is the first gene of the *srfa* operon
 C:Superfamily: surfactin synthetase; acetate-CoA ligase homology; acyl carrier protein h
 C:Keywords: antibiotic biosynthesis; carrier protein; phosphopantetheine; phosphoprotein
 F:507-950/Domain: acetate-CoA ligase homology <ACLI1>
 F:974-1042/Domain: acyl carrier protein homology <ACPI1>
 F:1043-1488/Domain: repeat <RPT1>
 F:1549-1993/Domain: acetate-CoA ligase homology <ACLI2>
 F:2011-2079/Domain: acyl carrier protein homology <ACP2>
 F:2080-2327/Domain: repeat <RPT2>
 F:2589-3025/Domain: acetate-CoA ligase homology <ACLI3>
 F:3047-3109/Domain: acyl carrier protein homology <ACP3>
 F:1006,2043,3074/Blinding site: phosphopantetheine (Ser) (covalent) #status predicted
 Query Match 4.7%; Score 95; DB 2; Length 3588;
 Best Local Similarity 19.7%; Pred. No. 45;
 Matches 63; Conservative 48; Mismatches 111; Indels 98; Gaps 13;
 14 VSOMALLFHMYHNTSSLMKAPREHNVLLSSWSSGSSFGQLFGCHPDVFTYLM 73

Db 990 VERAQIFDNFEETGHSLSAMTLTKIRK-----ETGIEIPQOFLFHEPTITAAE 1040
 Oy 74 PAMVMTFKOSTAMMLHMAVRDLIRAVFLCDMSVDAYMEPPGRSSILFQENSALC 133
 Db 1041 EADH-----RESNAF-----AYIEPRKEHNPILHSHSEHS 1073
 Oy 134 SAPA-----CDIIPDEIIT-----PRAHRCILCS 157
 Db 1074 SASSRRRESAIIHASSNSSECFRYSKAGARISGINPPIVETSFVLENSPPOKIH-VCV 1132
 Oy 158 QQPFVEVERAKRSYSHVLEKVFENLQSLYPLKDPSP-----NLHIV-----HLVRDP 208
 Db 1133 DENIEMTERKGNDEDAIMASFVTFDLARA-PLFRIGLGLGEENRMLPLEDMHLLTSDGV 1191
 Oy 209 AVFRSRRRTKGLMID--SRIVAGHOKLKKEDQPYVWQVYQOSOLEYTKTQISLPR 266
 Db 1192 SI-----GIMLELARIVYGEQLPDLRQKYAVWQ--SRQAEGYKKDQYWK 1240
 Oy 267 L--QERYLVRYEDLARAPV 284
 Db 1241 VFAGELPVLQSLDYPRPV 1260
 RESULT 8
 T13288
 mel-41 protein - fruit fly (*Drosophila melanogaster*)
 C:Species: *Drosophila melanogaster*
 C:Date: 13-Aug-1999 #sequence_rev1sion 13-Aug-1999 #text_change 17-Nov-2000
 C:Accession: T13288
 R:Harl, K.L.; Senterre, A.; Sekelsky, J.J.; McKim, K.S.; Boyd, J.B.; Hawley, R.S.
 Cell 62, 815-821, 1995
 A:Title: The mel-41 gene of *D. melanogaster* is a structural and functional homolog of
 A:Reference number: Z11072; MVID:95401271; PMID:7671309
 A:Accession: T13288
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2354 <HAP>
 A:Cross-references: EMBL:U34925; NID:g998351; PID:g998353; PIDN:AAC46881.1
 C:Genetics:
 A:Gene: mel-41
 A:Cross-references: FlyBase:Fgn0004367
 A:Introns: 650/3; 748/3; 2313/3
 C:Function:
 A:Description: Involved in cell cycle checkpoint and meiotic recombination
 Query Match 4.5%; Score 92.5; DB 2; Length 2354;
 Best Local Similarity 20.5%; Pred. No. 42;
 Matches 85; Conservative 52; Mismatches 147; Indels 131; Gaps 20;
 Oy 28 SHNRS---LSMKAQPERHNVLLSSWSSGSSFGQLFGCHPDVFTYLMPEAMHWMTFKO 84
 Db 1058 NVNFASPOHPLNLTISDDFAVLATLSLCRGYQF--QQNTKHYDSFSL----- 1102
 Oy 85 STAMMLHMAVRDLIRAVFLCDMSVFPAYMEPPGRSSILFQENSALCSAPACDIIPD 144
 Db 1103 -----AIQETLA---TCGIS-----PKKKKQVLMQ-----SLPA---RMQ 1133
 Oy 145 EIIPRAHRCILCSQPFVEVERAK--RSYSHVLEKVF-----NLQS 186
 Db 1134 LMEPMHSCYTCVHRSTQLQPLRPSHSNHYTEMALMARSRLDIPSSGKRLLSS 1193
 Oy 187 LYPPLK-----PSLNLHIV-----HLVRDRAVFRSRRRTKGLMIDSRIV 228
 Db 1194 YKPCIKRDSNMLSTFPYLLHALLECTEQRNHIQEEPMAYLQANEES-----SSSV 1246
 Oy 229 MCHQHEKLLKEDQPYVWQVYQOSOLEYTKTQISLPAQERYLVRYEDLARAPVQTS 288
 Db 1247 RGROELGAIKEN-----AFKQFESRRKYAGI-KPLASTLYSDRKEDSRVP----- 1291
 Oy 289 RMYEFGVLEFLPHLOTW-----VHNITGKMGDAHFTNAR---DALNVQAMRWSLP 339
 Db 1292 RLAGKLCALBLDFLQRLRMEMORIHGRSTG--CKPPTIDSNYRKTHIEFLNL-----IP 1343

OY 340 YEKSRRLQKACGDAMNLLGRIHVRSQ-----EQRNLLDLISTWTVPEQI 385
 ||| ||: | | |: | |:: | |:
Db 1344 KLVASRASYNCGEYARALSYLSYLEEGEDKSQRLLGEQTFIVEVGYGLRDPDSV 1398

RESULT 9
T16350
hypothetical protein F42G9.8 - *Caenorhabditis elegans*

A:Gene: CESP:F42G9.8
A:Introns: 28/2; 50/2; 113/2; 155/3; 220/3; 290/2
C:Superfamily: Caenorhabditis elegans hypothetical protein F42G9.8

Qy 52 RGSSEFVGOLFQHPDV---FYLMEPAHWMTFKOSTWMLHNAVRDLIRAVELCDMS 107
 |||::: |||
 Db 95 RSGTTLIRAILADPVRGCGETMLPSFTLWQGRND--VWNNST-----TOE 143
 |||::: |||

```
QY      167 ACRSHVVLKEVFENFNLQSLYPPLKDP SLNLIHVHVRDPRAVFRSREFTK ---GDLM 222Z
```

279 LARAIVAGNTRMYEVLGFLPLHLOTWVHNITRGKSGDHAFH-TNARDALNVQAWRWS 3377

Qy 338 LPY-EKVSRLQKACGDAMNLGY 359
: | : : ||
Db 312 DCFSEETLRKLDVAPFLGILGY 334

Query Match	4.5%	Score 91.5	DB 2	Length 338
Best Local Similarity	22.28%	Pred. No. 3.8		
Matches 52	Conservative 32	Mismatches 79	Indels 71	Gaps 11

```
OY      67 DVEYLMEPAHVMWTFKQSTAMMLMAVRDLIRAVFLCDSVFADAYNEPGPARQSLSFQM   126  
       | : | | | : | : |  
Db     59 DIVITYP-----KSGTNMLE-----IVCLIKD-----GDPNRI RSEPITWO-    96
```

```

QY      186 SLPLPKPSLNLIHVLVRDPRAVFRSRETKGDLMDISLIVMGQHEQLKKEQDQYV 245
          |      ::::|::|::|      |      |      |      |
Db      134 -----SSKAVIYYVGNPRDYYVS-----LYYSKIA-GQ---LKDPGTPDQF 172

```

RESULT 11
S11737
resistance protein Mx3, interferon-regulated - rat

Query Match	4.58;	Score 91;	DB 2;	Length 659;
Best Local Similarity	20.68;	Pred. No. 10;		
Matches 65; Conservative	50;	Mismatches 104;	Indels 96;	Gaps 16

```

QY      115 -----PGRROSSLFQWENSALCSAPACDILPQDEIIPRACHRLCS-QQPF--- 161
      |      : : : |      :      : : : |      : : : |
Db      354 ELQKYGADIPEDENKTLFLIEKINAF-NODITAYEGEEIVREKECRLLFTYLRKEPFLW 412

```

```
QY      201 VHLVRD-PRAYER-----SRERTKGLMIDRSIYNGQHOKLKKEDPYYVMQV 248
```

DB. 328 IICQDDIIR-----KALQR-----VREELAEELERKHGASNSAUSPNUJISSMD----- 3/2

QY 297 EFLPHLQTN---VHN 308

A:Residues: 1-338 <SAK>
A:Cross-references: GB:AF026072; NTD:g2570897; PIDN:AAC69918.1; PID:g2570898
C:Comment: This enzyme catalyzes the transfer of a sulfonate group from the active sulfid
P:
C:Superfamily: alcohol sulfotransferase
C:Keywords: transferase

0Y-----KOSTAMLMHMAVRDLIRAVFLCDMSVEVDAYMEBQPROSSLEFOW 126

Qy 325 RDALINVSQAMRWSLPEYKVSRLQKACGDAMNLGTRHVRSEQEOIRNLLDLL 376
| | :|::| | | | :|:

RESULT 15

C69679

polyketide synthase pksM - Bacillus subtilis

C/Species: Bacillus subtilis

C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 03-Nov-2000

C/Accession: C69679

R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A/Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holzapfel, S.; Hosono, S.; Hullo, M.F.

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

T.; Winters, P.; Wipet, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A/Authors: Yoshikawa, H.F.; Zunshein, E.; Yoshikawa, H.; Danchin, A.

A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A/Reference number: A69580; MUID:98044033; PMID:9384377

A/Accession: C69679

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-4273 <RUN>

A/Cross-references: GB:299113; GB:AL009126; NID:g2634090; PIDN:CAB13603.1; PID:g2634103

A/Experimental source: strain 168

C/Genetics:

A/Gene: pksM

C/Superfamily: Bacillus subtilis polyketide synthase pksM; 3-oxoacyl-[acyl-carrier-prote

C/Keywords: carrier protein

F:293-363/Domain: acyl carrier protein homology <ACPI>

F:414-829/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OASI>

F:2240-2732/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OASI>

F:3145-3320/Domain: short-chain alcohol dehydrogenase homology <SADI>

F:3550-3942/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OASI>

F:4138-4208/Domain: acyl carrier protein homology <ACP>

Query Match

Best Local Similarity 4.4%; Score 89.5; DB 2; Length 4273;

Matches 48; Conservative 34; Mismatches 87; Indels 41; Gaps 10;

```

QY      161 FEVEKACRSYSHVLEKVEFFNQLSLPLKDP-----LNLHIVHLVRDPRAYRSR 216
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      36 YQVQEGHYAVQDELK-----NLTFFPLADESDIALTIHVSREBGTWSTIIDGQ 90
QY      217 TKGDLMTDSRIYMGQHE-QRLKREDQPYVMQVIGOSQLETKTIGSLPRALQERYLLVR 275
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      91 QHGESLSDKR---QYETADMHRKEQYAFASI-----DLNQMKSTADRIINLDEIYEQCR 142
QY      276 YEDLARPAVQOT-SRME-----FVGLIEFLPLQOTWVHNIT--RGKMGCDHANHTN 323
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      143 SOELVHTGMKAKAQIYKAEAGAVIDLAVQOELRHSDAFLFHTLIDSGIGS----- 196
QY      324 ARDALNVSQAMRWSLPEKYS---RLQKAC 350
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      197 --SCLISDQTMVILPLIYESFSASERLQKC 224

```

Search completed: August 9, 2003, 16:44:57
 Job time : 16.9612 secs

THIS PAGE BLANK (USPTO)

RA MEDLINE-97305956; PubMed-9163423;
 RA Feilberg C.A., Fellay R., Broughton W.J., Rosenthal A.,
 RA Perret X.;
 RT "Molecular basis of symbiosis between Rhizobium and legumes.";
 CC Nature 387:394-401(1997).
 CC -1- FUNCTION: REQUIRED FOR THE FORMATION OF SULFATED NOD FACTOR.
 CC PROPOSED TO TRANSFER ACTIVATED SULFATE (PAPS) TO THE PICOSE
 CC OF THE NOD FACTOR.
 CC -1- SIMILARITY: LIMITED TO NODH AND TO C.ELEGANS F4269.8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Y09415; CAA70569.1; -;
 DR EMBL; AE000076; AAB91690.1; -;
 DR InterPro: IPR000863; Sulfoltransferase.
 DR Pfam: PF00685; Sulfoltransfer 1.
 KW Nucleation; Transferase; Plasmid.
 SQ SEQUENCE 419 AA; 46569 MW; 848C48E0416AAALF CRC64;
 Query Match 5.1%; Score 104.5; DB 1; Length 419;
 Best Local Similarity 21.5%; Pred. No. 0.19;
 Matches 78; Conservative 48; Mismatches 134; Indels 103; Gaps 18;
 QY 34 LSKAKQPERMNVILVLSWSSGSSFGQLFGQHPDVFYLMPEAMHVMFTKOSTAMLMHA 93
 DB 1 MSRDVSPPLPICFLCFLGIPRSGTTLAHLQHPDITAPRP-----WL----- 43
 QY 94 VRLIRAVFLCDMSVPFAYVPEPRRROSSLFQWENSALCSAPACDIIPODEIT----- 147
 DB 44 -----MLALEAFGRVDHHPA-----GASLIQVACEPLGRIDRISVRVF 84
 QY 148 -PRAHCRILCSQPFEEVERKACRSYSHVLEKEVFEMQLQSLYPLKPSINLHIVHVRD 206
 DB 85 ADAAYSGYLAAAKRRTIDKTPRW--NVID-----YLHSLYF--EAP-----HL-LLRN 130
 QY 207 PRAVFSRERTKGDMLDSRIYVNGHEQKLKEDQPYVYVQVTCQSQ-----LE 255
 DB 131 PYAIAASLKTGWCPFEVSEKCP-----PTVSQIALAVTGTPTAAVALA 174
 QY 256 IYTTISLPAKIDR-----YLVRYEDLARAPVAGTSRM-----YEEVGLEFLPHLTGW 306
 DB 175 LADLVGLPALAMQGRHRTQVRYERLVERPDEIQRLVAGLGYDPAGIVFAGVEQT-- 232
 QY 307 HNTTRGKMGD-HAFHTNARDALNVSQAMRMSLPYEKVSRLQKACG-DAMNLLGY-----R 360
 DB 233 -EYLRSSPFGDRRLKKAANDNSV-ETWTELTIEEMQVITDVLGADLVLELGYEQSLQ 290
 QY 361 HVR 363
 DB 291 HAR 293
 RESULT 4
 SRF1_BACSU
 ID SRF1_BACSU STANDARD; PRT; 3588 AA.
 AC P27206;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Surfactin synthetase subunit 1.
 GN SRF4A OR SRF41 OR SRF4.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RT

RA MEDLINE-93181186; PubMed-8441623;
 RA Funa S., Fujishima Y., Corbell N., D'Souza C., Nakano M.M.,
 RA Zuber P., Yamane K.;
 RT "Nucleotide sequence of 5' portion of srfA that contains the region
 RT required for competence establishment in Bacillus subtilis.";
 CC Nucleic Acids Res. 21:93-97(1993).
 CC [12]
 CC SEQUENCE FROM N.A.
 CC STRAIN=168 / JH642;
 CC MEDLINE-93360813; PubMed-8355609;
 CC Cosmina P., Rodriguez F., de Ferra F., Grandi G., Perego M.,
 CC Venema G., van Sinderen D.;
 CC "Sequence and analysis of the genetic locus responsible for surfactin
 CC synthesis in Bacillus subtilis.";
 CC Mol. Microbiol. 8:821-831(1993).
 CC [13]
 CC SEQUENCE FROM N.A.
 CC STRAIN=168;
 CC MEDLINE-97124189; PubMed-8969502;
 CC Yamane K., Kumano M., Kurita K.;
 CC "The 25 degrees-36 degrees region of the Bacillus subtilis
 CC chromosome: determination of the sequence of a 146 kb segment and
 CC identification of 113 genes.";
 CC Microbiology 142:3047-3056(1996).
 CC [14]
 CC SEQUENCE FROM N.A.
 CC STRAIN=168;
 CC MEDLINE-98044033; PubMed-9384377;
 CC Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 CC Azevedo V., Bertolo M.G., Bessieres P., Bolotin A., Borchert S.,
 CC Borries R., Boursier L., Brans A., Braun M., Brigelli S.C., Bron S.,
 CC Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 CC Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 CC Denlioz F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 CC Enlian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 CC Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
 CC Ghm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 CC Giuseppe G., Guy B.J., Haga K., Haleb J., Harwood C.R., Henaut A.,
 CC Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 CC Joris B., Karamata D., Kasahara Y., Klaert-Blanchard M., Klein C.,
 CC Kodayashl Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
 CC Kurita K., Lapilus A., Lardinois S., Lauber J., Lazarevic V.,
 CC Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 CC Medina N., Meliado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 CC Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 CC Patro V., Pohl T.M., Portetelle D., Porroli S., Prescott A.M.,
 CC Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 CC Rieger M., Rivolta C., Roche B., Roche B., Rose M., Sadate Y.,
 CC Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 CC Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Solido B.,
 CC Socotkin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 CC Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 CC Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
 CC Viari A., Wambuit R., Wedler E., Wedler H., Weitzenecker T.,
 CC Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 CC Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 CC "The complete genome sequence of the Gram-positive bacterium Bacillus
 CC subtilis.";
 CC Nature 390:249-256(1997).
 CC [15]
 CC SEQUENCE OF 1-460 FROM N.A.
 CC MEDLINE-91154134; PubMed-1847909;
 CC Nakano M.M., Magnuson R., Myers A.M., Curry J., Grossman A.D.,
 CC Zuber P.;
 CC "srfA is an operon required for surfactin production, competence
 CC development, and efficient sporulation in Bacillus subtilis.";
 CC J. Bacteriol. 173:1770-1778(1991).
 CC [16]
 CC SEQUENCE OF 1-64 FROM N.A.
 CC STRAIN=168;
 CC MEDLINE-95219080; PubMed-7704255;
 CC Fujishima Y., Yamane K.;
 CC "A 10 kb nucleotide sequence at the 5' flanking region (32 degrees)

RT of srfaA of the Bacillus subtilis chromosome. ;
 RL Microbiology 141:277-279(1995).
 RP SEQUENCE OF 1-38 FROM N.A.
 RX MEDLINE-91358326; PubMed-1715856;
 RA Nakano M.M., Xia L., Zuber P.;
 RT "transcription initiation region of the srfa operon, which is
 controlled by the comp-coma signal transduction system in Bacillus
 subtilis." ;
 RL J. Bacteriol. 173:5487-5493(1991).
 CC -1- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME ABLE TO
 CC ACTIVATE AND POLYMERIZE THE AMINO ACIDS LEU, GLU, ASP AND VAL.
 CC ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
 CC -1- CORRECTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
 CC -1- PATHWAY: Cyclic peptide antibiotic surfactin biosynthesis.
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 CC FAMILY.
 CC -1- SIMILARITY: Contains 3 acyl carrier domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC -----
 CC EMBL: D13262; BAA02522.1; -
 CC EMBL: X70356; CAA49816.1; -
 CC EMBL: D50453; BAA08982.1; -
 CC EMBL: Z99105; CAB12142.1; -
 CC EMBL: M59939; AAA22815.1; -
 CC EMBL: D30762; BAA21034.1; -
 CC EMBL: M64702; AAA22816.1; -
 CC PIR: I40485; I40485.
 CC HSP: P14687; JAMU.
 CC Subtilist: BG10168; srfaA.
 CC InterPro: IPR000873; AMP-bind.
 CC InterPro: IPR001242; Condensatn.
 CC InterPro: IPR006163; PP_bind.
 CC InterPro: IPR006162; Pantne_attach.
 CC Pfam: PF00501; AMP-binding; 3.
 CC Pfam: PF00668; Condensation; 4.
 CC Pfam: PF00550; pp-binding; 3.
 CC PRINTS: PRO0154; AMPBINDING.
 CC PROSITE: PS00012; PHOSPHOPANTETHEINE; 2.
 CC PROSITE: PS00455; AMP BINDING; 3.
 CC PROSITE: PS00075; ACP DOMAIN; 3.
 CC Llgase: Antibiotic biosynthesis; Phosphopantetheine; Sporulation;
 CC Multifunctional enzyme; Repeat; Complete proteome.
 CC REPEAT ? 1047 DOMAIN 1 (GLU-ACTIVATING).
 CC REPEAT ? 2084 DOMAIN 2 (LEU-ACTIVATING).
 CC DOMAIN ? 3115 DOMAIN 3 (D-LEU-ACTIVATING).
 CC DOMAIN 976 1043 ACYL CARRIER (ACP) 1.
 CC DOMAIN 2013 2080 ACYL CARRIER (ACP) 2.
 CC DOMAIN 3044 3110 ACYL CARRIER (ACP) 3.
 CC BINDING 1006 1006 PHOSPHOPANTETHEINE (POTENTIAL).
 CC BINDING 2043 2043 PHOSPHOPANTETHEINE (POTENTIAL).
 CC BINDING 3074 3074 PHOSPHOPANTETHEINE (POTENTIAL).
 CC BINDING 146 146 M -> I (IN REF. 1 AND 5).
 CC CONFLICT 151 151 V -> L (IN REF. 1 AND 5).
 CC CONFLICT 165 165 G -> A (IN REF. 1).
 CC CONFLICT 281 281 Q -> T (IN REF. 1).
 CC CONFLICT 460 460 D -> T (IN REF. 5).
 CC CONFLICT 540 540 A -> P (IN REF. 1).
 CC CONFLICT 562 562 Y -> I (IN REF. 1).
 CC CONFLICT 639 640 PT -> GS (IN REF. 1).
 CC CONFLICT 644 644 P -> R (IN REF. 1).
 CC CONFLICT 649 649 FLIP -> LLA (IN REF. 1).
 CC CONFLICT 1026 1026 Q -> L (IN REF. 1).
 CC CONFLICT 1065 1114 HHSSEHSSASSRMRSEAIHSSNSSEGFYSKAGARISG
 CC IMPRTVTE -> SLAQQRTYVLSQFQEDGCGYNNPAAAIL
 CC EGPDIQKLEAFQGLIRHESLR (IN REF. 1).

FT CONFLICT 1130 1131 VC -> DS (IN REF. 1).
 FT CONFLICT 1162 1162 A -> R (IN REF. 1).
 FT CONFLICT 1456 1456 L -> V (IN REF. 1).
 FT CONFLICT 1848 1848 T -> S (IN REF. 1).
 FT CONFLICT 1892 1892 T -> D (IN REF. 1).
 FT CONFLICT 1909 1909 A -> R (IN REF. 1).
 FT CONFLICT 1971 1978 PAVEIQMD -> LRCSKWT (IN REF. 1).
 FT CONFLICT 2050 2050 P -> L (IN REF. 1).
 FT CONFLICT 2216 2216 Q -> H (IN REF. 1).
 FT CONFLICT 2263 2264 HR -> QQ (IN REF. 1).
 FT CONFLICT 2289 2289 L -> V (IN REF. 1).
 FT CONFLICT 2347 2347 C -> S (IN REF. 1).
 FT CONFLICT 2426 2426 S -> R (IN REF. 1).
 FT CONFLICT 2609 2610 AV -> RC (IN REF. 1).
 FT CONFLICT 2684 2686 MISSING (IN REF. 1).
 FT CONFLICT 2756 2756 S -> T (IN REF. 1).
 FT CONFLICT 2897 2898 TA -> SP (IN REF. 1).
 FT CONFLICT 3026 3026 N -> P (IN REF. 1).
 FT CONFLICT 3097 3097 N -> F (IN REF. 1).
 FT CONFLICT 3272 3272 S -> A (IN REF. 1).
 FT CONFLICT 3317 3317 S -> R (IN REF. 1).
 FT CONFLICT 3452 3452 S -> Y (IN REF. 1).
 FT CONFLICT 3484 3496 HOMSHRPFTTSH -> DEMSDAGLTTRE (IN REF. 1).
 FT CONFLICT 3500 3501 PH -> GQ (IN REF. 1).
 FT CONFLICT 3508 3508 E -> R (IN REF. 1).
 SQ SEQUENCE 3588 AA; 402072 MW; BC02FB15D7F1FDB CRC64;
 Query Match 4.7%; Score 95; DB 1; Length 3588;
 Best Local Similarity 19.7%; Pred. No. 19; Indels 98; Gaps 13;
 Matches 63; Conservative 48; Mismatches 111;
 QY 14 VSOMATLALFFHYNISNLSKAPERNHVLVLSMRSGSFVQLFGQHPDYVLM 73
 DB 990 VERAGIFDNFEFFEGHSLKAMTLTKIHK-----ETGIEIPQGFLEHPTITLALME 1040
 QY 74 PAMHWMTFKOSTAMLMHNAVRLINAVFLCDSVDAVMEQPPROSSLFQWENSRAIC 133
 DB 1041 EADH-----RESKAF-----AVIEPAKQEHYPLHHSSEHS 1073
 QY 134 SAPA-----CDIIPODEII-----PRAHCRLLCS 157
 DB 1074 SASRRRESAIIHSSNSGCFRYSKAGARISGINTPPRIEYSFVLENSTPROKIH-VCV 1132
 QY 158 QQFEVEYERACRSYSHVLEKVFEMIQSLYPLNDPSL-----NLHIV-----HLVRDR 208
 DB 1133 DNIENIERGSGDEAIIASVFTFDLANA-PLFRIGLGLLENRMHLFDMHHLISDGV 1191
 QY 209 AVFRSEKRTKGLMID--SRIVMGQHEQKKEKEDPYIYQVYICQSLITYKTIOSLPRA 266
 DB 1192 SI-----GIMLEELARIYKGEQLPDLRIQKDYAVWQ--SRQPAEGYKRDQAYWKE 1240
 QY 267 L--QERYLLVREYEDLARAVP 284
 DB 1241 VTAGELPVLQILSDIPRPV 1260
 RESULT 5
 TPRS CAEEL STANDARD; PRT; 359 AA.
 AC Q20351;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Putative protein-tyrosine sulfotransferase (EC 2.8.2.20)
 DE (Tyrosylprotein sulfotransferase) (TPST).
 GN F42G9.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.


```
CC -1- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.  
CC -1- SIMILARITY: Contains 1 SH3 domain.  
CC -1- SIMILARITY: Contains 37 spectrin repeats.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isdb.ch/announce/  
CC or send an email to licensel@isb-sib.ch).  
-----  
DR EMBL; AB029290; BAA83821.1; -  
DR EMBL; AF111968; AAF06360.1; -  
DR EMBL; AF325341; AAL39000.1; -  
DR EMBL; AF325330; AAL39000.1; JOINED.  
DR EMBL; AF325331; AAL39000.1; JOINED.  
DR EMBL; AF325332; AAL39000.1; JOINED.  
DR EMBL; AF325333; AAL39000.1; JOINED.  
DR EMBL; AF325334; AAL39000.1; JOINED.  
DR EMBL; AF325335; AAL39000.1; JOINED.  
DR EMBL; AF325336; AAL39000.1; JOINED.  
DR EMBL; AF325339; AAL39000.1; JOINED.  
DR EMBL; AF325340; AAL39000.1; JOINED.  
DR EMBL; AB033077; BAA86565.1; -  
DR EMBL; AL137853; CAC15920.1; -  
DR EMBL; AB007934; BAA32310.2; -  
DR PIR; T00079; T00079.  
DR HSSP; O01082; IBRK.  
DR Genew; HGNC:13664; MACF1.  
DR GO; GO:0005856; C:cytoskeleton; NAS.  
DR GO; GO:0003780; F:actin cross-linking activity; NAS.  
DR GO; GO:0005509; F:calcium ion binding activity; NAS.  
DR GO; GO:0008017; F:microtubule binding activity; NAS.  
DR InterPro; IPRO01589; Actbind actln.  
DR InterPro; IPRO01715; Calponin-like.  
DR InterPro; IPRO02048; EF-hand.  
DR InterPro; IPRO03108; GAS2.  
DR InterPro; IPRO01452; SH3.  
DR InterPro; IPRO02017; Spectrln.  
DR Pfam; PF00307; CH; 2.  
DR Pfam; PF00036; ehand; 2.  
DR Pfam; PF02187; GAS2; 1.  
DR Pfam; PF00435; spectrin; 27.  
DR ProDom; PD000012; EF-hand; 1.  
DR SMART; SM00033; CH; 2.  
DR SMART; SM00054; EFh; 2.  
DR SMART; SM00243; GAS2; 1.  
DR SMART; SM00150; SPEC; 36.  
DR PROSITE; PS00019; ACTININ_1; 1.  
DR PROSITE; PS00020; ACTININ_2; FALSE_NEG.  
DR PROSITE; PS00021; CH; 2.  
DR PROSITE; PS00018; EF_HAND; 2.  
DR PROSITE; PSS0002; SH3; FALSE_NEG.  
KW Actin-binding; Cytoskeleton; Calcium; Calcium-binding; Repeat;  
KW SH3 domain; Coiled coil; Alternative splicing.  
FT DOMAIN 1 295  
FT DOMAIN 78 181  
FT DOMAIN 194 295  
FT DOMAIN 243 265  
FT DOMAIN 477 529  
FT DOMAIN 722 751  
FT DOMAIN 816 843  
FT DOMAIN 1013 1118  
FT DOMAIN 1164 1191  
FT DOMAIN 1199 1690  
FT DOMAIN 1780 1843  
FT DOMAIN 1975 2005  
FT DOMAIN 2039 2312  
FT DOMAIN 2385 2417  
FT DOMAIN 2544 2695  
FT DOMAIN 2760 2838
```

FT	DOMAIN	2911	3001	COILED COIL (POTENTIAL).	
FT	DOMAIN	3130	3164	COILED COIL (POTENTIAL).	
FT	DOMAIN	3244	3277	COILED COIL (POTENTIAL).	
FT	DOMAIN	3418	3482	COILED COIL (POTENTIAL).	
FT	DOMAIN	3596	3666	COILED COIL (POTENTIAL).	
FT	DOMAIN	3786	3806	COILED COIL (POTENTIAL).	
FT	DOMAIN	3852	3931	COILED COIL (POTENTIAL).	
FT	DOMAIN	3967	3987	COILED COIL (POTENTIAL).	
FT	DOMAIN	4084	4218	COILED COIL (POTENTIAL).	
FT	DOMAIN	4343	4378	COILED COIL (POTENTIAL).	
FT	DOMAIN	4408	4437	COILED COIL (POTENTIAL).	
FT	DOMAIN	4468	4498	COILED COIL (POTENTIAL).	
FT	DOMAIN	4907	4935	COILED COIL (POTENTIAL).	
FT	DOMAIN	5044	5067	COILED COIL (POTENTIAL).	
FT	REPEAT	314	355	SPECTRIN 1.	
FT	REPEAT	591	623	SPECTRIN 2.	
FT	REPEAT	680	784	SPECTRIN 3.	
FT	REPEAT	766	800	SPECTRIN 4.	
FT	DOMAIN	871	923	SH3.	
FT	REPEAT	1250	1272	SPECTRIN 5.	
FT	REPEAT	1287	1342	SPECTRIN 6.	
FT	REPEAT	1455	1534	SPECTRIN 7.	
FT	REPEAT	1547	1659	SPECTRIN 8.	
FT	REPEAT	1815	1891	SPECTRIN 9.	
FT	REPEAT	1932	2042	SPECTRIN 10.	
FT	REPEAT	2260	2280	SPECTRIN 11.	
FT	REPEAT	2372	2395	SPECTRIN 12.	
FT	REPEAT	2398	2507	SPECTRIN 13.	
FT	REPEAT	2510	2618	SPECTRIN 14.	
FT	REPEAT	2621	2728	SPECTRIN 15.	
FT	REPEAT	2731	2838	SPECTRIN 16.	
FT	REPEAT	2841	2945	SPECTRIN 17.	
FT	REPEAT	2987	3024	SPECTRIN 18.	
FT	REPEAT	3136	3163	SPECTRIN 19.	
FT	REPEAT	3187	3274	SPECTRIN 20.	
FT	REPEAT	3277	3383	SPECTRIN 21.	
FT	REPEAT	3386	3492	SPECTRIN 22.	
FT	REPEAT	3495	3601	SPECTRIN 23.	
FT	REPEAT	3604	3673	SPECTRIN 24.	
FT	REPEAT	3713	3819	SPECTRIN 25.	
FT	REPEAT	3832	3927	SPECTRIN 26.	
FT	REPEAT	3982	4043	SPECTRIN 27.	
FT	REPEAT	4046	4152	SPECTRIN 28.	
FT	REPEAT	4155	4262	SPECTRIN 29.	
Query Match					
Best Local Similarity		4.5%;	Score 91.5;	DB 1; Length 5430.	
Matches		88; Conservative	66; Mismatches	128; Indels 115; Gaps 24;	
Oy	2	LLPKMKMLLVLVSOMATLAFPHNYSINISLSMKAPERNHVLVLSWHSGSFVGOL	61		
Db	1752	LLSQONFILTKQSOA-----FLDGHGHLT-----PERQML-----OQKDEL	1792		
Oy	62	EGOHDPVEYLMEPAHWMTFKOSTAWMLHM-AVRDLIRAVFLCDMSVFADYMERGPRRO	120		
Db	1793	KEQXS-----TSLAQSFAELKQVOTLDLOEK-FLDHKKEFESMLE---RSE	1835		
Oy	121	SSLQWENSRALCSAPACDIIIPDEIIPRACRILCSQPF--EVEKACSYSH-	173		
Db	1836	KEL---ENMKGGSSP-----ETTP-----SLTKRGSSSEDT-----SHKGDLR	1873		
Oy	174	-VVLKEVFFNLSQSLYPLLKDPSSLNLHIVLVPRAVFRSRETRKGDMLTDSR-----	226		
Db	1874	FVITSGQVLDMENSFKEGRKPS---ELGNLVKD-----KIKDATERITALHSKCTRGS	1925		
Oy	227	---IVMGQHECKLKKEDDPYTVWVOYICSOLE--YKTIQSLPKALDERVLLVRYEDLAR	281		
Db	1926	HLNMLLGYHOFONSADSLQAMMOA-CEANVEKLTSLDVADPGVLOEO-----	1973		
Oy	282	APVAVOTSMYEVFGLEFLP--HLQTPWYNI--TGKKGGDIAFHTNARDA-LNYSQAWRM	336		
Db	1974	--LWITTKOLOBELAEHOVPEKLVKAVANDIMEISGAPRDRRHVOETTDLSLHSPQSLSY	2031		

OY 337 SLPYKVSRLQKACGDAMNLGVRHVRSEOEORILL 373
 DB 2032 SLA-ERSSLQKAIAS-----QSVESLESL 2058

RESULT 7
 ID MAC4_HUMAN STANDARD; PRT; 5938 AA.
 AC 096PK2; 08WXY1;
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Microtubule-actin crosslinking factor 1, isoform 4.
 GN MACF1 OR ACF1 OR ABP620 OR KIAA0465 OR KIA1251.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.
 RX MEDLINE-21833812; PubMed-11845288;
 RA Gong T.-W.L., Bestiri C.G., Lomax M.I.;
 RT "MACF1 gene structure: a hybrid of plectin and dystrophin.";
 RL Mamm. Genome 12:852-861(2001).
 CC -1- FUNCTION: May play a role in cross-linking cytoskeletal proteins
 by binding intermediate filaments to the N-terminal plectin
 repeats and microtubules to the C-terminal.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 Event-Alternative splicing; Named isoforms=4;
 CC Name=4;
 CC IsoId=096PK2-1; Sequence-Displayed;
 CC Name=1;
 CC IsoId=090PN3-1; Sequence-External;
 CC Name=2;
 CC IsoId=090PN3-2; Sequence-External;
 CC Name=3;
 CC IsoId=090PN3-3; Sequence-External;
 CC -1- TISSUE SPECIFICITY: Expressed in heart, lung, pituitary and
 placenta, not found in brain, kidney, liver, pancreas or skeletal
 muscle.
 CC -1- SIMILARITY: Belongs to the plectin or cytolinker family.
 CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
 CC -1- SIMILARITY: Contains 19 plectin repeats.
 CC -1- SIMILARITY: Contains 32 spectrin repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF317696; AAL09459.1; -;
 DR EMBL: AF325341; AAL38997.1; -;
 DR EMBL: AF325333; AAL38997.1; JOINED.
 DR EMBL: AF325334; AAL38997.1; JOINED.
 DR EMBL: AF325335; AAL38997.1; JOINED.
 DR EMBL: AF325336; AAL38997.1; JOINED.
 DR EMBL: AF325339; AAL38997.1; JOINED.
 DR EMBL: AF325340; AAL38997.1; JOINED.
 DR GO: GO:0008017; F-microtubule binding; ISS.
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR003108; GAS2.
 DR InterPro: IPR001101; Plectin_repeat.
 DR Pfam: PF000036; ehand; 2.
 DR Pfam: PF02187; GAS2; 1.
 DR Pfam: PF00681; Plectin; 11.
 DR Pfam: PF00435; Spectrin; 26.
 DR ProDom: PD000012; EF-hand; 1.
 DR SMART: SM00054; Efn; 2.

DR SMART: SM00243; GAS2; 1.
 DR SMART: SM00250; PLEC; 19.
 DR SMART: SM00150; SPEC; 32.
 DR PROSITE: PS00018; EF_HAND; 2.
 DR Cytokeleton; Calcium; Calcium-binding; Repeat; Coiled coil;
 KW Alternative splicing.
 FT DOMAIN 1830 1936 COILED COIL (POTENTIAL).
 FT DOMAIN 2001 2192 COILED COIL (POTENTIAL).
 FT DOMAIN 2282 2345 COILED COIL (POTENTIAL).
 FT DOMAIN 2477 2507 COILED COIL (POTENTIAL).
 FT DOMAIN 2541 2654 COILED COIL (POTENTIAL).
 FT DOMAIN 2686 2814 COILED COIL (POTENTIAL).
 FT DOMAIN 2887 2919 COILED COIL (POTENTIAL).
 FT DOMAIN 3046 3197 COILED COIL (POTENTIAL).
 FT DOMAIN 3262 3503 COILED COIL (POTENTIAL).
 FT DOMAIN 3532 3666 COILED COIL (POTENTIAL).
 FT DOMAIN 3746 3779 COILED COIL (POTENTIAL).
 FT DOMAIN 3920 3984 COILED COIL (POTENTIAL).
 FT DOMAIN 4098 4168 COILED COIL (POTENTIAL).
 FT DOMAIN 4288 4308 COILED COIL (POTENTIAL).
 FT DOMAIN 4354 4386 COILED COIL (POTENTIAL).
 FT DOMAIN 4397 4433 COILED COIL (POTENTIAL).
 FT DOMAIN 4469 4489 COILED COIL (POTENTIAL).
 FT DOMAIN 4586 4720 COILED COIL (POTENTIAL).
 FT DOMAIN 4845 4880 COILED COIL (POTENTIAL).
 FT DOMAIN 4910 4939 COILED COIL (POTENTIAL).
 FT DOMAIN 4970 5000 COILED COIL (POTENTIAL).
 FT DOMAIN 5409 5437 COILED COIL (POTENTIAL).
 FT DOMAIN 5546 5569 COILED COIL (POTENTIAL).
 FT CA_BIND 5598 5610 EF_HAND 1 (POTENTIAL).
 FT CA_BIND 5634 5646 EF_HAND 2 (POTENTIAL).
 FT REPEAT 12 49 PLECTIN 1.
 FT REPEAT 53 88 PLECTIN 2.
 FT REPEAT 89 126 PLECTIN 3.
 FT REPEAT 130 164 PLECTIN 4.
 FT REPEAT 166 202 PLECTIN 5.
 FT REPEAT 203 240 PLECTIN 6.
 FT REPEAT 243 278 PLECTIN 7.
 FT REPEAT 279 316 PLECTIN 8.
 FT REPEAT 318 354 PLECTIN 9.
 FT REPEAT 725 762 PLECTIN 10.
 FT REPEAT 763 800 PLECTIN 11.
 FT REPEAT 801 836 PLECTIN 12.
 FT REPEAT 839 878 PLECTIN 13.
 FT REPEAT 897 934 PLECTIN 14.
 FT REPEAT 935 972 PLECTIN 15.
 FT REPEAT 975 1011 PLECTIN 16.
 FT REPEAT 1012 1049 PLECTIN 17.
 FT REPEAT 1121 1158 PLECTIN 18.
 FT REPEAT 1159 1196 PLECTIN 19.
 FT REPEAT 1925 2032 PLECTIN 1.
 FT REPEAT 2052 2160 PLECTIN 2.
 FT REPEAT 2211 2313 PLECTIN 3.
 FT REPEAT 2320 2430 PLECTIN 4.
 FT REPEAT 2437 2543 PLECTIN 5.
 FT REPEAT 2550 2652 PLECTIN 6.
 FT REPEAT 2659 2758 PLECTIN 7.
 FT REPEAT 2765 2896 PLECTIN 8.
 FT REPEAT 2903 3008 PLECTIN 9.
 FT REPEAT 3015 3119 PLECTIN 10.
 FT REPEAT 3126 3229 PLECTIN 11.
 FT REPEAT 3236 3339 PLECTIN 12.
 FT REPEAT 3346 3446 PLECTIN 13.
 FT REPEAT 3453 3555 PLECTIN 14.
 FT REPEAT 3562 3664 PLECTIN 15.
 FT REPEAT 3671 3775 PLECTIN 16.
 FT REPEAT 3782 3884 PLECTIN 17.
 FT REPEAT 3891 3993 PLECTIN 18.
 FT REPEAT 4000 4102 PLECTIN 19.
 FT REPEAT 4109 4211 PLECTIN 20.
 FT REPEAT 4218 4320 PLECTIN 21.
 FT REPEAT 4327 4428 PLECTIN 22.
 FT REPEAT 4438 4544 PLECTIN 23.

```

FT REPEAT 4551 4653 SPECTRIN 24.
FT REPEAT 4660 4763 SPECTRIN 25.
FT REPEAT 4770 4872 SPECTRIN 26.
FT REPEAT 4879 4982 SPECTRIN 27.
FT REPEAT 4989 5091 SPECTRIN 28.
FT REPEAT 5098 5201 SPECTRIN 29.
FT REPEAT 5208 5309 SPECTRIN 30.
FT REPEAT 5316 5418 SPECTRIN 31.
FT REPEAT 5425 5525 SPECTRIN 32.
FT CONFLICT 1712 1712 P -> S (IN REF. 1: AAL38997).
SQ SEQUENCE 5938 AA; 670134 MW; B8784112752DA004 CRC64;

Query Match 4.5%; Score 91.5; DB 1; Length 5938;
Best Local Similarity 22.2%; Pred. No. 74;
Matches 88; Conservative 66; Mismatches 128; Indels 115; Gaps 24;

QY 2 ILPRKMKLLFLVQSQALALFEHMYSHNSSLKKAQERHNVLVLSMRGSSRYGL 61
DB 2254 LLSQONFILTATQSHQA---FLDQGHNL-----PEQOML-----QOKLSEL 2294
QY 62 FGOHDPVYLLMPAWHVMTEFKOSTAMMLHM-AVBDLRAVFLCDMSYFADAMEPGRRO 120
DB 2295 KEQYS-----TSLAQSEALKEQVOTLQDELQK-FLDQHKFESWLE---RSE 2337
QY 121 SLPFGWENSRALCSAPACDIIPODELIPRAHRLCSQOPF---EYVERACRSYSH----- 173
DB 2338 KEL---ENMHKGGSSP-----ETLP-----SLKROGSFSEVDYI-----SHKGLR 2375
QY 174 -VLKEVFEFNLOSYPPLKDPSLNLIHVLVRDPAVRSRERTKGLMIDSR----- 226
DB 2376 FTISGQAVLDMENEFKSGKERS---EIGNLVKD-----KLMDATERIALMSKCRIGS 2427
QY 227 ---YMGQHEQKLKEDQPYVMQYICQSLQ---IKTIQSLPKALQERYLLVREEDLAR 281
DB 2428 HLMMLLGGYHOPQNSDAQMMAQ--CEANVEKLLSDYVAPSDPGVLOEQ----- 2475
QY 282 AVNAQTSRMYEVEGLEFLP--HLOTWVNI--TRCKMGMDHAFHTNARA--LVNSQAMRW 336
DB 2476 ---LATTKQIQEELAHQVFEKLVQVARDIMEESGPADRHVHOETTDLSLHQSLSY 2533
QY 337 SLPEKVSRLQKACGDAMNLGVRHVRSEQEOBNLL 373
DB 2534 SLA-ERSILLQKAIQS-----QSVQESLESLL 2560

RESULT 8
339F_HUMAN
ID 239F_HUMAN STANDARD: PRT; 294 AA.
AC 015777;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Fetal brain protein 239 (239FB).
GN C11ORF8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=95080775; PubMed=752372;
RA Schwartz F., Neve R., Eisenman R., Gessler M., Bruns G.;
RT "A WAGR region gene between PAX-6 and FSHB expressed in fetal brain.";
RL Hum. Genet. 94:658-664(1994).
RN 12
RP SEQUENCE FROM N.A.
RX MEDLINE=96115606; PubMed=8666403;
RA Schwartz F., Eisenman R., Knoll J., Gessler M., Bruns G.;
RT "cDNA sequence, genomic organization, and evolutionary conservation
of a novel gene from the WAGR region.";
RL Genomics 29:526-532(1995).
RN 13
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=Colon, Kidney, and Stomach;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Pelngold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Scheffer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Bouffard G.G.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Miliady S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley A.C., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Smalls D.E.,
RA Butlerfield I.S.N., Krzywinski M.I., Skalska U.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN FETAL BRAIN.
CC -1- SIMILARITY: BELONGS TO THE UPF0046 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.sdb.ch).
CC
DR EMBL: U57911; AAC50564.1;
DR EMBL: BC031582; AAH31582.1;
DR Genew; HGNC:1180; C11orf8.
DR MIM: 600911;
DR GO: GO:0007399; P-neurogenesis; TAS.
DR InterPro: IPR004843; M-peptidase.
DR Pfam: PF00149; Metallophos; 1.
SQ SEQUENCE 294 AA; 33360 MW; 43B2BCDAA1BFD1F0 CRC64;

Query Match 4.5%; Score 91; DB 1; Length 294;
Best Local Similarity 20.2%; Pred. No. 17;
Matches 53; Conservative 30; Mismatches 89; Indels 90; Gaps 9;

QY 139 DIIPODELIPRAHRLCSQOPFEYVERACRSYSHVL-----KEYREFNLSL 187
DB 45 DIPIDTPKPAHTRVCISDHSKRIDGOMYGDILHTGFTGLSPSEVKKFN----- 100
QY 188 YELLKDPSSLNLIHVLVRDPAVRSRERTKGLMIDSRIV--GOHQKLKKE-----D 240
DB 101 -----DWLGNLPYEVKIYVAGNHELFDEKFEFADLVK 132
QY 241 QPYVMQVYCQSOLEIKTIQSLPKALQERYLLVREEDARPAVQTSRM-----YEVGL 296
DB 133 QDYTRPPSVSKLPEDFDVQSL-----LTNSIYQDSVYTKGFRITGA 177
QY 297 EFLPHLOTGVNHTTRKMGMDHAFHTNARDALVNSQAMWSLPYEKVSRLQKACGDAMNL 356
DB 178 PWTPMNGMNGFNLPRGQSLD-----KWNLIPEIDILM--THGPP----- 216
QY 357 LGYRHVRSQEQEQRNLLDLILST 378
DB 217 LGFRDWPVKELQRYVGCVELLNT 238

RESULT 9
MX3_RAT
ID MX3_RAT STANDARD: PRT; 659 AA.
AC P18590;
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last sequence update)

```



```

DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Interferon-induced GTP-binding protein Mx3.
GN Mx3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91055598; PubMed=2173790;
RA Meier E., Kunz G., Haller O., Arnhelter H.;
RT "Activity of rat Mx proteins against a rabdovirus.";
RL J. Virol. 64:6263-6269(1990).
CC -1- FUNCTION: DOES NOT SHOW ACTIVITY AGAINST INFLUENZA VIRUS OR VSV;
CC ALTHOUGH IT ONLY DIFFERS FROM Mx2 BY 8 POSITIONS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- INDUCTION: By Interferons.
CC -1- SIMILARITY: BELONGS TO THE DYNAMIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52713; CA36937.1; -
DR PIR: S11737; S11737.
DR InterPro: IPR001401; Dynamln.
DR InterPro: IPR000375; Dynamln_central.
DR InterPro: IPR003130; GED.
DR Pfam: PF00350; dynamln_1.
DR Pfam: PF01031; dynamln_2; 1.
DR Pfam: PF02212; GED; 1.
DR PRINTS: PR00195; DYNAMIN.
DR SMART: SM00053; DYNC; 1.
DR SMART: SM00302; GED; 1.
DR PROSITE: PS00410; DYNAMIN; 1.
KW Interferon induction; GTP-binding; Multigene family.
FT NP_BIND 75 82 GTP (POTENTIAL).
FT NP_BIND 176 180 GTP (POTENTIAL).
FT NP_BIND 245 248 GTP (POTENTIAL).
FT SEQUENCE 659 AA; 74951 MW; F55D63283BC865B3 CRC64;
SQ
Query Match 4.58; Score 91; DB 1; Length 659;
Best Local Similarity 20.68; Pred. No. 4.9;
Matches 65; Conservative 50; Mismatches 104; Indels 96; Gaps 16;
QY 62 FGHPDVEYLMPEAMHVMFQKSTAWMLHMAVVDLI---RAVELCDMSVFDAVME--- 114
DB 301 FKEHPORALLLEDG-----KATVPCLAEKLTWELLISHICKSLPLENQIKESHQSTSE 353
QY 115 -----PGPRROSLFQWENSRLACSPACDIIPDEIIPRAHCRLLCS-QQPF--- 161
DB 354 ELQKYGADIPEDENKELFLEKINAF-NODITAIVEGEELVREKECKLFKLRKEFLM 412
QY 162 -EVEKACRSYSHVLEKREFNQ-----SLYPLKDSPLNHI 200
DB 413 SEELERNFQKSDALYEVYTFEFGYGRGLPGFVNKTFENILRQIKTLEPPAME-- 470
QY 201 VHLVRD-PRAVFR-----SRRTKGLMIDSRIVMGHOKLKREDOYVYVMOV 248
DB 471 LHKTEIVRAFTVSEKNESEFNNLHRTYTSKIEDIRL---EOTEKAKIRLHFQWEO 527
QY 249 ICQSQLEIYKTIQSLPALOEYLLVREDI-----ARAPYAQTSRYEYVGL 296
DB 528 ILYQODIYR-----KALOK-----VREEAEEERKHKGRSAQSPMLQTSMD----- 572
QY 297 EFLPHLOTW---VHN 308
DB 573 EIFOHLNAYROEAIN 587

```

```

RESULT 10
C7D8_SOYBN
ID C7D8_SOYBN STANDARD: PRT: 504 AA.
AC 081974;
DR 15-DEC-1998 (Rel. 37, Created)
DR 15-DEC-1998 (Rel. 37, Last sequence update)
DR 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome P450 71D8 (EC 1.14.-.-) (P450 CYP7).
GN CYP71D8.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
RX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Harosoy 63;
RX MEDLINE=98311068; PubMed=9648734;
RA Schopfer C.R., Ebel J.;
RT "Identification of elicitor-induced cytochrome P450s of soybean
RL (Glycine max L.) using differential display of mRNA.";
RL Mol. Gen. Genet. 258:315-322(1998).
CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y10493; CA71517.1; -
DR PIR: T07120; T07120.
DR HSSP: P14779; IJPZ.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Heme.
FT METAL 444 444 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT SEQUENCE 504 AA; 57579 MW; 86E4FAEEC2F2A8F CRC64;
SQ
Query Match 4.48; Score 89.5; DB 1; Length 504;
Best Local Similarity 19.98; Pred. No. 4.7;
Matches 70; Conservative 45; Mismatches 108; Indels 129; Gaps 16;
QY 68 VFYLMPEAMHVMFQKSTAWMLHMAVVDLI---RAVELCDMSVFDAVMEPGPRROSLFQWE 127
DB 14 VFLLIH-WLY-KTYKQKSHKL-----PGPMR----- 39
QY 128 NSRALCSAPACDIIPDEIIPRAHCRLLCSQOPEVEKACRSYS---HVLKVEYFENL 184
DB 40 -----LP---IGNLHQLAALASLPDQALQVLRKGPRLHQLGISTLVV 83
QY 185 QSLYPLKDSPLNHIYLVADPRPAVFRFSRRTKGLMIDSRIVMGHOKLKREDOY 244
DB 84 SS--PKAMEMKMKTHDHFVQRPQ-----LLAPQFMYGATDIAPADGYWR 129
QY 245 VMQVYCQSLIYKTIQSLPALOE---RYLLVREYEDLARAPVAQTSRYEYVGL----- 296
DB 130 QIRKICTELLLSARVQSFSHIRQDENKKLQSIHSSAGSPIDLSGRKFSLGTIVSRAA 189
QY 297 -----EFL-----PHLOTWVHNITRKGKMGDAFHTNA----- 324
DB 190 FGENDDODEFMSVYRAKITWTGFEVDMPSPILKP-LHLITROKAKVEHV-HORADKIL 247
QY 325 RDALNVSGAMRMNSLPYEKVSRLQKACGAMNLGLRIHVRSPQEDRNILLDL 376
DB 248 EDILRKHM-----EKRTVRKKGNG-----SEAEQEDLVALL 279

```

RESULT 11
PKSM_BACSU STANDARD; PRT: 4273 AA.
AC P40872; 031781.
DT 01-FEB-1996 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative polyketide synthase pksm.
GN PKSM OR PKSY.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertorello M.G., Bessières P., Bolotin A., Borchert S., Boriss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A., Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T., Enlitan K.D., Errington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N., Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G., Giuseppe G., Guy B.J., Haga K., Hachez J., Harwood C.R., Henaut A., Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapdus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetlelle D., Portolillo S., Prescott A.M., Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F., Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B., Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Yamakoshi A., Yanaka T., Terpstra P., Tognoni A., Totoso V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Wetzemeyer T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A., "The complete genome sequence of the Gram-positive bacterium Bacillus subtilis." Nature 390:249-256(1997).
RT Nature 390:249-256(1997).
RL [2]
RN SEQUENCE OF 1-1763 FROM N.A.
RP STRAIN=168 / PBI1424;
RX Tognoni A., Grandi G.;
RA Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN SECONDARY METABOLISM.
CC -1- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTHETHEINES (POTENTIAL).
CC -1- SIMILARITY: Contains 4 acyl carrier domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC
CC EMBL; 299113; CAB13603.1; -
CC EMBL; 235133; CA84505.1; -
CC PIR; C69679; C69679.
CC Subtilist; BG10931; PKSM.
CC InterPro; IPR002198; ADH_short.
CC InterPro; IPR000794; Ketoacyl-synt.

DR InterPro; IPR001601; Methyltransf.
DR InterPro; IPR006163; pp_bind.
DR InterPro; IPR006162; Pantine-attach.
DR InterPro; IPR000051; SAM_bind.
DR Pfam; PF00106; adh_short; 1.
DR Pfam; PF02801; ketoacyl-synt; 3.
DR Pfam; PF00550; ketoacyl-synt; 4.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 2.
DR PROSITE; PS00606; B_KETOACYL-SYNTHASE; 2.
DR PROSITE; PS0075; ACP_DOMAIN; 4.
KM Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
FT DOMAIN 295
FT DOMAIN 364
FT DOMAIN 396
FT DOMAIN 2190
FT DOMAIN 2258
FT DOMAIN 2322
FT DOMAIN 3532
FT DOMAIN 3410
FT DOMAIN 4140
FT BINDING 327
FT BINDING 327
FT ACT_SITE 2222
FT ACT_SITE 2476
FT BINDING 3446
FT ACT_SITE 3690
FT BINDING 4172
FT CONFLICT 103
FT CONFLICT 276
FT CONFLICT 289
SQ SEQUENCE 4273 AA; 477459 MW; 3BBFC1A250AB54 CMC64;
Query Match 4.48; Score 89.5; DB 1; Length 4273;
Best Local Similarity 22.98; Pred. No. 72;
Matches 48; Conservative 34; Mismatches 87; Indels 41; Gaps 10;

161 FEYVERACRSYSHVKEVFNLSQYPLKXPS---LNLHIVLVDPRAVFSRER 216
36 YQVFEHGYATQLEIK-----NLTFYPLADESDIALTLTVSREESTWSTITDGC 90
217 TKGDLMDISRIWQHE-OKLKEDQPYVMQYICOSLEIYVTLSPKALQERYLV 275
91 QHESLSDKR---QYETADMHKREQTAFNEST---DLNOMSTADRLINDEIYEQOR 142
276 YEDLAAPVPAQT-SRATY-----FYGLEPLHLQVWVNIIT--RKGMDAHFHN 323
143 SELVHTGMMKAGQYIEAKGAVIDLVAQGEALRHSDFLPHPTLIDSGIGS----- 196
324 ARDALNVSOAMRWSLPEKVS---RLQKAC 350
197 ---SCLISDOTMTPLPYESFSASERLQKGC 224
Db

RESULT 12
FAST FLACH STANDARD; PRT: 320 AA.
AC P52837;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Flavonol 4'-sulfotransferase (EC 2.8.2.-) (F4-S7).
OS Flaveria chloroefolia.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; OC Asteridae; campanulids; Asterales; Asteraceae; Tageteae; OC Flaveria.
OX NCBI_TaxID=4228;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92159034; PubMed=1741382;
RA Varin L., Deluca V., Ibrahim R.K., Brisson N.;
RT "Molecular characterization of two plant flavonol sulfotransferases." J. Mol. Natl. Acad. Sci. U.S.A. 89:1286-1290(1992).
RN [2]

RP PAPS-BINDING SITE.
 RX MEDLINE=95279378; PubMed-7759495;
 RA Varin L., Marsolais F., Brisson N.;
 RT "Chimeric flavonol sulfotransferases define a domain responsible for
 RT substrate and position specificities."
 RL J. Biol. Chem. 270:12498-12502(1995).
 CC -1- FUNCTION: TRANSFERS SULFATE GROUP INTO FLAVONOLIN POSITION 4'.
 CC MAY PLAY A ROLE IN AUXIN TRANSPORT.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- TISSUE SPECIFICITY: HIGHEST IN SHOOT TIPS AND LOWEST IN MATURE
 CC LEAVES AND ROOTS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE PLANT SULFOTRANSFERASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M64136; AAA3343.1; -
 DR HSSP: P50224; 1CJM;
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 DR ProDom: PD001218; Sulfotransferase; 1.
 KM Transferrase.
 FT BINDING 148 163 PAPS-BINDING SITE.
 SQ SEQUENCE 320 AA; 37255 MW; 6B2659F6CCBC0 CRC64;
 Query Match 4.4%; Score 89; DB 1; Length 320;
 Best Local Similarity 20.8%; Pred. No. 2.9;
 Matches 56; Conservative 30; Mismatches 83; Indels 100; Gaps 12;
 QY 54 GSSEVQIQREGQR-DVFIYMEPRAMHWMTFKOSTAMHLMAVNDLIAVFLCMSVDFAY 112
 Db 47 GALLAQSFKAPDDVFLCSYP-----KSGTITWKLALVAIVTR-----EKDEF 91
 QY 113 MEP-----GPRROSSLFQ-WENSALCSAPACDIIPODEIIPRACHRLCSQOFE 162
 Db 92 TSPILNINIPNCIPYEIKDKLTIENNNSCFTR-----NATHPIYH 133
 QY 163 VVEKACRSYSHVVLKEVRFNLSLYPLKDPISLNIHVLVDPRAVFRSRETKDLM 222
 Db 134 VLKRS-----IL-----ALNCMKVIYIRIKDY----- 156
 QY 223 IDSRIVMGHEOKLKKEDOPYYMVQVICSQLEIYKIQSLP-----KALQER-- 270
 Db 157 IVSFYHFGREITIKLPEDAPFE-----EAFDEFYHGISQFGPYWHLGLGYMKASLERPE 210
 QY 271 -YLLVRYEDLARAIPAQTSRMVEFVGLF 298
 Db 211 VILFKYEDYKDKPTSVNKLAEFIGYF 239
 RESULT 13
 MPPA_RAT
 ID MPPA_RAT STANDARD: PRT; 524 AA.
 AC P20069;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Mitochondrial processing peptidase alpha subunit, mitochondrial
 DE precursor (EC 3.4.24.64) (Alpha-MPP) (P-55).
 GN INPSE OR PMPCA OR MPPA.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RX MEDLINE=91045920; PubMed=2236012;

RA Kleiber J., Kalousek F., Swaroop M., Rosenberg L.E.;
 RT "The general mitochondrial matrix processing protease from rat liver:
 RT structural characterization of the catalytic subunit."
 RL Proc. Natl. Acad. Sci. U.S.A. 87:7978-7982(1990).
 CC -1- FUNCTION: THE MITOCHONDRIAL PROCESSING PROTEASE (MPP-I) CLEAVES
 CC PRESEQUENCES FROM MITOCHONDRIAL PROTEIN PRECURSORS. MOST MPP-I
 CC CLEAVAGE SITES FOLLOW AN ARGININE AT POSITION -2.
 CC -1- CATALYTIC ACTIVITY: Release of N-terminal transit peptides from
 CC precursor proteins imported into the mitochondrion, typically with
 CC Arg in position P2.
 CC -1- SUBUNIT: Heterodimer of alpha and beta subunits.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
 CC -1- CAUTION: Does not seem to have protease activity as it lacks the
 CC zinc-binding site.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M57728; AAA41632.1; -
 DR PIR: A36205; A36205.
 DR MEROPS: M16.971; -
 DR InterPro: IPR001431; Peptidase_M16.
 DR Pfam: PF00675; Peptidase_M16; 1.
 DR Pfam: PF05193; Peptidase_M16_C; 1.
 DR PROSITE: PS00143; INSULINASE; 1.
 KM Hydroxylase; Metalloprotease; Mitochondrion; Transit peptide.
 FT TRANSIT 1 32
 FT CHAIN 33 524
 FT SUBUNIT.
 SQ SEQUENCE 524 AA; 58607 MW; BBF08FBC9F09DB2 CRC64;
 Query Match 4.4%; Score 89; DB 1; Length 524;
 Best Local Similarity 21.5%; Pred. No. 5.4;
 Matches 70; Conservative 46; Mismatches 104; Indels 106; Gaps 19;
 QY 70 YLM--EPAM-----HWMTFQ-----STAMHLMAVNDLIAVFLCMSVDFAYMEPG 116
 Db 269 YLLGVCPAMGAPVAMVLTQWHSSTRGSSRW-----RE-----TCOMSA---LRP 311
 QY 117 PRROSSLFQWENSALCSAPACDIIPODEIIPRACHRLC-----SQPFVEYE 165
 Db 312 PRROSSHI-YGAREL-----LLLEDFIPFAVLNMWGGGGSFAGRGKMF-- 360
 QY 166 KACRSYSHVVLKEVRFNLSLYPLKDPISLNIHVLVDPRAVFRSRETKDLMIDS 225
 Db 361 ---RLYINVLNRHHMMWMTNATSYHSHYED--TGLLCIHASADPRQV-----REWEIITKE 410
 QY 226 RIYWGHEOKLKKEDOPYYMVQVICSQLEIYKIQSLPALKOERYLLVRYEDLARAIPA 285
 Db 411 FILMGRTVDVLELERAFTQLSMLMN-----LESRP-----VIFEDVGROYLA 454
 QY 286 QTSRMVEFVGLFEPH-LQTVVNHITRGKMGDAFTNARDALNVSQARWMSLPIYEKVS 344
 Db 455 THSRK-----LPHELCTILIRNV-----KPEDIKRVASKMLRGKP----- 488
 QY 345 RLQKACGDANMLGVRHVRSEQEQRN 370
 Db 489 -AVALGLDLDLPYEHIOAALSBD 513
 RESULT 14
 CAPP_STRPY
 ID CAPP_STRPY STANDARD: PRT; 920 AA.
 AC Q9A0U7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

OY 278 DLAPAPVAQTSRMKEFVGLFPLPLQTW---VHNIT--RGKGM 315
Db 596 NNLYGLTNYTEGKRAVTVGVONL-----WSNVHEVSTKRHKGM 633

Search completed: August 9, 2003, 16:43:05
Job time : 12.4755 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:40:04 ; Search time 34.4109 Seconds
(without alignments)
2894.677 Million cell updates/sec

Title: US-09-645-078-2
Perfect score: 2038
Sequence: 1 MLPKRMKLLFLVSQMAIL.....EQRNLLDLSTWTFEQIH 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rviro:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	2038	100.0	386	4	Q9Y5R3
2	2035	99.9	386	4	Q8NCG5
3	1963	96.3	370	4	Q8IV46
4	1515.5	74.4	388	11	Q9RI11
5	1513.5	74.3	388	11	Q9WUE5
6	1019	50.0	395	4	Q9GZX3
7	1017.5	49.9	395	11	Q9GUP4
8	1008	49.5	411	4	Q9GZS9
9	627	30.8	484	11	Q9EP78
10	624	30.6	484	11	Q9EP78
11	606.5	29.8	486	4	Q75667
12	605.5	29.7	486	4	Q9NS84
13	598.5	29.4	530	11	Q88276
14	591.5	29.0	483	4	Q9UEB5
15	591.5	29.0	530	4	Q9Y4C5
16	559	27.4	411	11	Q9EBC0

17	549	26.9	411	4	Q43916	043916 homo sapien
18	504.5	24.8	472	11	Q88199	088199 mus musculu
19	482	23.7	479	4	Q75099	075099 homo sapien
20	481.5	23.6	474	11	Q9GRL2	09GRL2 ratu
21	423.5	20.8	441	13	Q93403	093403 torped cal
22	264	13.0	486	5	Q9VMC3	09VMC3 drosophila
23	215	10.5	120	11	Q9DOK5	09DOK5 mus musculu
24	210	10.3	363	5	Q9VMC4	09VMC4 drosophila
25	199	9.8	363	5	Q8MZD1	08MZD1 drosophila
26	175	8.6	119	6	Q957R8	0957R8 oryctolagus
27	168	8.2	307	16	Q92VY4	092VY4 thizobium m
28	124.5	6.1	274	16	Q8PRA0	08PRA0 xanthomonas
29	115.5	5.7	303	16	Q93JY6	093JY6 streptomyce
30	111	5.4	655	6	Q9BGC0	09BGC0 macaca fasc
31	108.5	5.3	1222	4	Q8IZU8	08IZU8 homo sapien
32	107	5.3	1586	4	Q9P2D3	09P2D3 homo sapien
33	100.5	4.9	747	4	Q96QD2	096QD2 homo sapien
34	100.5	4.9	762	17	Q8Q0G7	08Q0G7 methanosarc
35	100.5	4.9	775	4	Q9C011	09C011 homo sapien
36	100	4.9	593	10	Q8VZE0	08VZE0 arabidopsis
37	97.5	4.8	285	2	Q8KRX0	08KRX0 xanthomonas
38	96.5	4.7	743	4	Q9H0K2	09H0K2 homo sapien
39	96	4.7	2454	3	Q9UV56	09UV56 emericella
40	96	4.7	2454	3	Q9UV56	09UV56 emericella
41	95	4.7	2454	3	Q9UV56	09UV56 emericella
42	93.5	4.6	529	9	Q9V490	09V490 drosophila
43	93.5	4.6	1201	11	Q8C7A8	08C7A8 mus musculu
44	93	4.6	637	11	Q8C773	08C773 mus musculu
45	92.5	4.5	2334	5	Q24135	024135 drosophila

ALIGNMENTS

RESULT 1
ID Q9Y5R3 PRELIMINARY; PRT; 386 AA.
AC Q9Y5R3;
DT 01-NOV-1999 (TREMUR, 12, Created)
DT 01-NOV-1999 (TREMUR, 12, Last sequence update)
DT 01-MAR-2003 (TREMUR, 23, Last annotation update)
DE N-acetylglucosaminase 6-O-sulfotransferase (L-selectin ligand
DE sulfotransferase GSP-3).
GN GSP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Tonsil;
RX MEDLINE=99264336; PubMed=10330415;
RA Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,
RA Huang C.C., Kahanagi R., Rosen S.D., Hemmerich S.;
RT "Sulfotransferases of two specificities function in the reconstruction
RT of high endothelial cell ligands for L-selectin.";
RL J. Cell Biol. 145:899-910(1999).
[2]
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Tonsil;
RX MEDLINE=99264336; PubMed=10330415;
RA Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,
RA Huang C.C., Kahanagi R., Rosen S.D., Hemmerich S.;
RT "Sulfotransferases of two specificities function in the reconstruction
RT of high endothelial cell ligands for L-selectin.";
RL J. Cell Biol. 145:899-910(1999).
[3]
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Tonsil;
RX MEDLINE=11332592; PubMed=11439191;
RA Yeh J.-C., Hiraoka N., Petryniak B., Nakayama J., Elites L.G.,
RA Rakuda D., Hindsgrau O., March J.D., Lowe J.B., Rakuda M.;
RT "Novel sulfated lymphocyte homing receptors and their control by a
RT core extension beta1,3-N-acetylglucosaminyltransferase.";

RL Cell 105:957-969(2001).
 DR EMBL: AF131325; AAD3015.1; -
 DR EMBL: AF280088; AAG48246.1; -
 DR EMBL: AF149783; AAK48417.1; -
 DR InterPro: IPR001092; HLH_baslc.
 DR InterPro: IPR000863; Sulfoltransferase.
 DR Pfam: PF00685; Sulfoltransferase.
 DR ProSite: PS00038; HLH_1; 1.
 DR Lectin; Selectin; Transferase.
 SW SEQUENCE 386 AA; 45133 MW; 0C3BB402241743A CRC64;

Query Match 100.0%; Score 2038; DB 4; Length 386;
 Best Local Similarity 100.0%; Pred. No. 7,4e-184;
 Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLTPKMKLLFLVSOAMALALFFHMYSHNISLSKKAQPERMHVLYLSWSSGSSFFVQ 60
 Db 1 MLTPKMKLLFLVSOAMALALFFHMYSHNISLSKKAQPERMHVLYLSWSSGSSFFVQ 60
 QY 61 LFGQHPDVYLYMPANHVMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQ 120
 Db 61 LFGQHPDVYLYMPANHVMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQ 120
 QY 121 SSLEFQENSRLCSAPACDIIPDEIIPRAHCRLLCSQPFVEVEKACSYSHVYLKEVR 180
 Db 121 SSLEFQENSRLCSAPACDIIPDEIIPRAHCRLLCSQPFVEVEKACSYSHVYLKEVR 180
 QY 181 FFMQSLYPLKDPISNLHIVLRDPAVFSRERTKGDLMIDSRIVGQHEQKLKED 240
 Db 181 FFMQSLYPLKDPISNLHIVLRDPAVFSRERTKGDLMIDSRIVGQHEQKLKED 240
 QY 241 OPIYVMOVICOSQLETKTQSLPKALOERYLLVREDLARAFAVQTSRYEFGLEFLP 300
 Db 241 OPIYVMOVICOSQLETKTQSLPKALOERYLLVREDLARAFAVQTSRYEFGLEFLP 300
 QY 301 HQTWVHNTTRGKMGDHAFTNARDALNVSQAMRSLPYEKYSRLQKACGDMNLGTR 360
 Db 301 HQTWVHNTTRGKMGDHAFTNARDALNVSQAMRSLPYEKYSRLQKACGDMNLGTR 360
 QY 361 HVSEOEORNLIDLSTVTPQIH 386
 Db 361 HVSEOEORNLIDLSTVTPQIH 386

RESULT 2

QY 08NCG5 PRELIMINARY; PRT; 386 AA.
 Db 08NCG5;
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Hypothetical protein FLJ90265.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
 RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahara K.,
 RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
 RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK074746; BAC1177.1; -
 DR Gene: HGNC:1972; CHST4.
 KW Hypothetical protein.
 SQ SEQUENCE 386 AA; 45160 MW; 861869348319E42A CRC64;

Query Match 99.9%; Score 2035; DB 4; Length 386;
 Best Local Similarity 99.7%; Pred. No. 1,4e-183;
 Matches 385; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLTPKMKLLFLVSOAMALALFFHMYSHNISLSKKAQPERMHVLYLSWSSGSSFFVQ 60
 Db 1 MLTPKMKLLFLVSOAMALALFFHMYSHNISLSKKAQPERMHVLYLSWSSGSSFFVQ 60
 QY 61 LFGQHPDVYLYMPANHVMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQ 120
 Db 61 LFGQHPDVYLYMPANHVMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQ 120
 QY 121 SSLEFQENSRLCSAPACDIIPDEIIPRAHCRLLCSQPFVEVEKACSYSHVYLKEVR 180
 Db 121 SSLEFQENSRLCSAPACDIIPDEIIPRAHCRLLCSQPFVEVEKACSYSHVYLKEVR 180
 QY 181 FFMQSLYPLKDPISNLHIVLRDPAVFSRERTKGDLMIDSRIVGQHEQKLKED 240
 Db 181 FFMQSLYPLKDPISNLHIVLRDPAVFSRERTKGDLMIDSRIVGQHEQKLKED 240
 QY 241 OPIYVMOVICOSQLETKTQSLPKALOERYLLVREDLARAFAVQTSRYEFGLEFLP 300
 Db 241 OPIYVMOVICOSQLETKTQSLPKALOERYLLVREDLARAFAVQTSRYEFGLEFLP 300
 QY 301 HQTWVHNTTRGKMGDHAFTNARDALNVSQAMRSLPYEKYSRLQKACGDMNLGTR 360
 Db 301 HQTWVHNTTRGKMGDHAFTNARDALNVSQAMRSLPYEKYSRLQKACGDMNLGTR 360
 QY 361 HVSEOEORNLIDLSTVTPQIH 386
 Db 361 HVSEOEORNLIDLSTVTPQIH 386

RESULT 3

QY 08IV46 PRELIMINARY; PRT; 370 AA.
 Db 08IV46;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Similar to carboxylate (N-acetylglucosamine 6-O) sulfoltransferase
 4.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tissue-Brain;
 RA Strausberg R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC035282; AAH35282.1; -
 KW Transferase.
 SQ SEQUENCE 370 AA; 43249 MW; EA8D76EA4E73C625 CRC64;

Query Match 96.3%; Score 1963; DB 4; Length 370;
 Best Local Similarity 100.0%; Pred. No. 8,4e-177;
 Matches 370; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 MALALFFHMYSHNISLSKKAQPERMHVLYLSWSSGSSFFVQLEFGQHPDVYLYMPAN 76
 Db 1 MALALFFHMYSHNISLSKKAQPERMHVLYLSWSSGSSFFVQLEFGQHPDVYLYMPAN 76
 QY 77 HMYMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQSSLEFQENSRLCSAP 136
 Db 77 HMYMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQSSLEFQENSRLCSAP 136
 QY 137 ACIDIIPDEIIPRAHCRLLCSQPFVEVEKACSYSHVYLKEVREFNLSQSLPLKDPST 196
 Db 137 ACIDIIPDEIIPRAHCRLLCSQPFVEVEKACSYSHVYLKEVREFNLSQSLPLKDPST 196
 QY 197 NLHIVLRDPAVFSRERTKGDLMIDSRIVGQHEQKLKEDQPIYVMOVICOSQLEI 256
 Db 197 NLHIVLRDPAVFSRERTKGDLMIDSRIVGQHEQKLKEDQPIYVMOVICOSQLEI 256
 QY 257 YKTIOSLPKALOERYLLVREDLARAFAVQTSRYEFGLEFLPQHTWVHNTTRGKMG 316
 Db 257 YKTIOSLPKALOERYLLVREDLARAFAVQTSRYEFGLEFLPQHTWVHNTTRGKMG 316


```

Db      241 YKTQSLPKALQERLRYEDLARAPYAQTSRMEFEVGLFELHQLTQWVHNITRGKGMG 300
Oy      317 DHAFHTNARDALNTSQAWRMSLPEKYSRLQKAGDANMLGYRHVSEQORNLDDL 376
Db      301 DHAFHTNARDALNTSQAWRMSLPEKYSRLQKAGDANMLGYRHVSEQORNLDDL 360
Oy      377 STWTVEQI 386
Db      361 STWTVEQI 370

```

RESULT 4

```

O9R111 PRELIMINARY: PRT: 388 AA.
ID 09R111
AC 09R111
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE L-selectin ligand sulfotransferase.
CN CHST4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99361934; PubMed=10435581;
Hiroaka N., Petyrniak B., Nakayama J., Tsuboi S., Suzuki M., Yeh J.C.,
Izawa D., Tanaka T., Miyasaka M., Lowe J.B., Fukuda M.;
RT "A novel, high endothelial venule-specific sulfotransferase expresses
6-sulfatase 1 Lewis(x), an L-selectin ligand displayed by CD34.";
RL EMBL: AF109155; AAD45579.1;
DR MGD: MGI:1349479; Chst4.
DR InterPro: IPR000863; Sulfotransferase.
DR Pfam: PF00685; Sulfotransferase.
DR Lectin: Selectin; Transferrase.
SQ SEQUENCE 388 AA; 44636 MW; 6D5371AFB684AEE CRC64;

```

```

Query Match 74.4%; Score 1515.5; DB 11; Length 388;
Best Local Similarity 72.8%; Pred. No. 1.5e-134;
Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;

Oy      1 MLPRKMKLLFLVSQMAITLFFHMYSHNITSLSMKQAPR-MHVLVLSMRSGSSPVG 59
Db      1 MMLLKGRLMFLFSQVIVLALFTIMSVHR--HLSQRESRPPVHVLVLSMRSGSSPVG 58
Oy      60 QLFQGHDPVFLMEPAHVMWMTFKOSTAMMLHMAVRDLIRAVFLCDMSVPFAYMEPPRR 119
Db      59 QLFQGHDPVFLMEPAHVMWMTFTSTAMKLMHMAVRDLIRAVFLCDMSVPFAYMPPRR 118
Oy      120 QSSLFQWENSALCSAPACDIIPODEIPRAHCRLLCSQPFVEVERKCRSYSHVYLKEV 179
Db      119 QSSLFQWENSALCSAPACDIEPPAHEISSPKHCKLCQGFDMVEKACRSHGFEVLKEV 178
Oy      180 RFFNLQSLYPLLPKPSLNLHIVHLVDRPRAVRSRERTKGLDMITVMGQHEOKLKE 239
Db      179 RFLSLQALYPLLPKPSLNLHIVHLVDRPRAVRSRERTTIELVDSHIVLQOHELTKEE 238
Oy      240 DQPYVMQVITQSOLEIKTQSLPKALQERYLLVRYEDLARAPYAQTSRMEFEVGLFEL 299
Db      239 DQPYVMQVITQSOLEIKTQSLPKALQERYLLVRYEDLARAPYAQTSRMEFEVGLFEL 298
Oy      299 PHLTQWVHNITRGKGMGHAFHTNARDALNTSQAWRMSLPEKYSRLQKAGDANMLGY 359
Db      299 PHLTQWVHNITRGKGMGHAFHTNARDALNTSQAWRMSLPEKYSRLQKAGDANMLGY 358
Oy      360 RHVSEQORNLDDLSTWTVEQI 385
Db      359 LQVRSQEQORNLDDLSTWTVEQI 384

```

```

RESULT 5
O9WDE5 PRELIMINARY: PRT: 388 AA.
ID 09WDE5
AC 09WDE5
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase (carbohydrate (Chondroitin
DE 6/keratan) sulfotransferase 4).
CN CHST4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99264336; PubMed=10330415;
Bistrup A., Bhakta S., Lee J.-K., Belov Y.Y., Gunn M.D., Zuo F.-R.,
Huang C.-C., Kannagi R., Rosen S.D., Hemmerlich S.;
RT "Sulfotransferases of two specificities function in the reconstruction
of high endothelial cell ligands for L-selectin.";
RL J. Cell Biol. 145:899-910(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=C57BL/6J; TISSUE=Testis;
Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
Fleischmann W., Gaasterland T., Gissi C., King B., Kochika H.,
Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
Schulz L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
Gustigich S., Hill D., Hofmann M., Hume D.A., Kanlaya M., Lee N.H.,
Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombereis P.,
Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Wilming L.,
Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AF131236; AAD33016.1;
DR MGD: MGI:1349479; Chst4.
DR InterPro: IPR000863; Sulfotransferase.
DR Pfam: PF00685; Sulfotransferase.
DR Transferrase.
SQ SEQUENCE 388 AA; 44694 MW; D1E9D796DF8574D CRC64;

```

```

Query Match 74.3%; Score 1513.5; DB 11; Length 388;
Best Local Similarity 72.8%; Pred. No. 2.4e-134;
Matches 281; Conservative 49; Mismatches 55; Indels 3; Gaps 2;

Oy      1 MLPRKMKLLFLVSQMAITLFFHMYSHNITSLSMKQAPR-MHVLVLSMRSGSSPVG 59
Db      1 MMLLKGRLMFLFSQVIVLALFTIMSVHR--HLSQRESRPPVHVLVLSMRSGSSPVG 58
Oy      60 QLFQGHDPVFLMEPAHVMWMTFKOSTAMMLHMAVRDLIRAVFLCDMSVPFAYMEPPRR 119
Db      59 QLFQGHDPVFLMEPAHVMWMTFTSTAMKLMHMAVRDLIRAVFLCDMSVPFAYMPPRR 118
Oy      120 QSSLFQWENSALCSAPACDIIPODEIPRAHCRLLCSQPFVEVERKCRSYSHVYLKEV 179
Db      119 QSSLFQWENSALCSAPACDIEPPAHEISSPKHCKLCQGFDMVEKACRSHGFEVLKEV 178
Oy      180 RFFNLQSLYPLLPKPSLNLHIVHLVDRPRAVRSRERTKGLDMITVMGQHEOKLKE 239
Db      179 RFLSLQALYPLLPKPSLNLHIVHLVDRPRAVRSRERTTIELVDSHIVLQOHELTKEE 238

```

DB 179 RELSLQALPLTLPDLPSLNLHVHLVDRPAVRFSRSHHTITELMVDSHIYLGHLTIKEE 238
 QY 240 DOPYVYMOVICOSQLEIYKTIOSLPALQERYLLVREEDLAPAPVQTSRMVEFVLEPL 299
 DB 239 DOPYAMKTIKSSQVDIVAKICTLPALQORYLFEEDLVAPLQOTRRLYFVGIDFL 298
 QY 300 PHLOTWVHNITRKGKGDHAFHTNARDALNVSAQMWSLPYEKVSRLOKACGDAMLLG 359
 DB 299 PHLOTWVHNITRKGKGDHAFHTNARDALNVSAQMWSLPYEKVSQLOACGEMDILG 358
 QY 360 RHVRSEOEQRNLIDLSTWVPEQI 385
 DB 359 LGVRSQEQEGNLSIDLSSSHILGQV 384

RESULT 6
 Q9GZX3 PRELIMINARY; PRT; 395 AA.

Q9GZX3: PRELIMINARY; PRT; 395 AA.
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 23, Last annotation update)
 DE N-acetylglucosamine 6-O-sulfotransferase GST-4beta (Corneal N-acetylglucosamine-6-O-sulfotransferase).
 GN GST4BETA OR GST6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R., Rosen S.D.;
 RT "Chromosomal Localization and Genomic Organization for the Galactose/N-Acetylglactosamine/N-Acetylglucosamine 6-O-Sulfotransferase Gene Family.";
 RL Glycobiology 0:0-0(2001).
 RN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20472330; PubMed=11017086;
 RA Akama T.O., Nishida K., Nakayama J., Watanabe H., Fujiwara T., Nakamura T., Doka A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S., Ozaki K., Kinoshita S., Shimomura Y., Tanigami A., Fukuda M.N.;
 RT "Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulfotransferase gene.";
 RL Nat. Genet. 26:237-241(2000).
 DR EMBL; AF219990; AAG26325.1; -;
 DR EMBL; AF219991; AAG26327.1; -;
 DR Gene; HGNC:6938; CHST6.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer; 1.
 KW TRANSFERASE.
 SQ SEQUENCE 395 AA; 44098 MW; 433CA60248A48P67 CRC64;

Query Match 50.0%; Score 1019; DB 4; Length 395;
 Best Local Similarity 54.5%; Pred. No. 1.1e-87;

Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

QY 2 LPLKKMLLFLVSOAAILALFFHMYSHNITSSMKAOPERMHVLYLSSMRSGSSVVGOL 61
 DB 14 LLAQFFLLFLVSRP-----GPPSPGGBARVAVLYLSSMRSGSSVVGOL 59
 QY 62 FQOHDPVFLMEPAMHVMTFKOSTAMLMHMAVRDLIRAVFLCDMSVFDAYMEPPRRQS 121
 DB 60 FQOHDPVFLMEPAMHVMTFKOSTAMLMHMAVRDLIRAVFLCDMSVFDAYMEPPRRQS 118
 QY 122 SIFQWENSRALCSAPACDIIPODEIIPRAHCRILCSQOPFEVEYKACRSYSHVYLKEVF 181
 DB 119 DLFQWAVSRALCSPPCASFPFGAISSEVCKPILCAROSFTLAREACRSYSHVYLKEVF 178
 QY 182 FNLQSLYPLLPKPSLNLHVHLVDRPAVRFSRSHHTITELMVDSDHIOHOKLKEQ 241

DB 179 FNLQSLYPLLPKPSLNLHVHLVDRPAVRFSRSHHTITELMVDSDHIOHOKLKEQ 237
 QY 242 PYVYMOVICOSQLEIYK--TIOSLPALQERYLLVREEDLAPAPVQTSRMVEFVLEPL 299
 DB 238 GLRVAREVCRSHRILAEATIKP--PPLGRVRYLVEEDLVAPLQOTRRLYFVGIDFL 296
 QY 300 PHLOTWVHNITRKGKGDHAFHTNARDALNVSAQMWSLPYEKVSRLOKACGDAMLLG 357
 DB 297 POLEAMHNITRSGSGARREAFKISSRNALNVSAQMRHALPFAKIRVOELCAGALQL 356
 QY 358 GYRVRSEOEQRNLIDLSTWVPEQI 386
 DB 357 GYRVRSEOEQRNLIDLSTWVPEQI 385

RESULT 7
 Q9QUP4 PRELIMINARY; PRT; 395 AA.

Q9QUP4: PRELIMINARY; PRT; 395 AA.
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE N-acetylglucosamine 6-O-sulfotransferase.
 GN CHST5 OR I-GLCNAC-6-ST.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Intestine;
 RX MEDLINE=99423499; PubMed=10491328;
 RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
 RT "Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-Sulfotransferase that is Highly Restricted to Intestinal Tissue.";
 RL Biochem. Biophys. Res. Commun. 263:543-549(1999).
 DR EMBL; AF176841; AAD56003.1; -;
 DR EMBL; AF176840; AAD56002.1; -;
 DR MGI; MGI:1931825; Chst5.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer; 1.
 KW TRANSFERASE.
 SQ SEQUENCE 395 AA; 44537 MW; 3FDF71B43ED383BE CRC64;

Query Match 49.9%; Score 1017.5; DB 11; Length 395;
 Best Local Similarity 56.4%; Pred. No. 1.6e-87;

Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;

QY 1 MLPLK--KKMLLFLVSOAAILALFFHMYSHNITSSMKAOPERMHVLYLSSMRSGSSV 58
 DB 1 MRLPRSSSTVMLSLMLWGTGLVF--LVSRQVPS--SPAGLGRVAVLYLSSMRSGSSV 56
 QY 59 GOLFGOHDPVFLMEPAMHVMTFKOSTAMLMHMAVRDLIRAVFLCDMSVFDAYMEPPRR 118
 DB 57 GOLFGOHDPVFLMEPAMHVMTFKOSTAMLMHMAVRDLIRAVFLCDMSVFDAYMEPPRR 115
 QY 119 ROSSLFQWENSRALCSAPACDIIPODEIIPRAHCRILCSQOPFEVEYKACRSYSHVYLKE 178
 DB 116 NISDLFQWAVSRALCSPPCEAFARGNISSEVCKPILCATRPFGLAOEACSSYSHVYLKE 175
 QY 179 VREFNLQSLYPLLPKPSLNLHVHLVDRPAVRFSRSHHTITELMVDSDHIOHOKLKEQ 238
 DB 176 VREFNLQSLYPLLPKPSLNLHVHLVDRPAVRFSRSHHTITELMVDSDHIOHOKLKEQ 234
 QY 239 EDQPYVYMOVICOSQLEIYK--TIOSLPALQERYLLVREEDLAPAPVQTSRMVEFVLE 297
 DB 235 ADPRLRVAVNEVCRSHRILAEATIKP--PPLGRVRYLVEEDLVAPLQOTRRLYFVGID 294
 QY 298 FLPHLOTWVHNITRKGKGDHAFHTNARDALNVSAQMWSLPYEKVSRLOKACGDAMLLG 355
 DB 295 LFPLOTWVHNITRKGKGDHAFHTNARDALNVSAQMRHALPFAKIRVOELCAGALQL 354
 QY 356 LGYRVRSEOEQRNLIDLSTWVPEQI 386

SQ SEQUENCE 484 AA: 54766 MW: 981955370/AB7193 CRC64;
 Query Match 30.8%; Score 627; DB 11; Length 484;
 Best Local Similarity 34.4%; Pred. No. 1.5e-50;
 Matches 137; Conservative 65; Mismatches 150; Indels 46; Gaps 7
 QY NISSLSMKAQPERMIVLVSSWRSGSSPFVGLFGCHPDVFLMEPAWHYMTFKOSTAM 89
 Db NISANGAVTQEKOHIVYHATWRTSSSEFGLFNHPDVFLEYEPWHLMOALYPGDAES 146
 QY LHMAYRDLIRAVFLCDMSVFDAVMPGPRQ-----SSLFQWENSRAALCSAPAC 139
 Db LOGALRDLRLSFLRCDFSVLRLYAQGDGGERAPDSANLTMTLFRMRTNKVYCSPLCP 206
 QY IIPDQ-----IIPRAHCILCSQDPEVEYKCRSTSHYLYKEVRFNIOSTLYPLIKDS 195
 Db AAPRADVGVGEDKACSESTCPSPVSLRALEACRKPVVIVKIDVRLDLGVLPILLRDPG 266
 QY LNLHIVHLVPRPRAVERSRERTKGLMTDS-----RTVM-----GQHE 233
 Db LNLKVVQLFPRDRAVHNSRLSRQGLRSTIOVLTFRQGDHFHRYLHLHGVDARPGQA 326
 QY QKLKREDDPY---VMQVICSQSEIYKTIQSLSPKALQERYLVAVEDIARAPAQTSR 289
 Db RALPSAPRADEFLLTSALFVIGCAWLRDLFTTGAPAWMLRRRLRLRYEDLVNQPOQLRR 386
 QY MTEPVGLEFLHLQWVHNITRGKMG-DHAFHTARDAALNVSQAMRSLPEYKVSRLQ 348
 Db LLRSGLTLTALAALDAFAFNRMGASVYAGDRPHLSARARAVAHAMRELTSEQVRYET 446
 QY ACGDAMNLTGYRHVRSQEOORNL-----LDLSTWTV 381
 Db ACAPAMRLIATPRSGDEDRKTVRGCEPLETAKAMVAY 484
 RESULT 10
 Q99NBO PRELIMINARY: PRT: 484 AA.
 AC Q99NBO:
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Chondroitin 6-sulfotransferase-2.
 GN CHST7 OR MC6ST-2.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA TISSUE=Mammary gland;
 RC Kitagawa H., Uyama T., Sugahara K.;
 RT "Cloning and Expression of Mouse Chondroitin 6-sulfotransferase-2";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB046929; BAB40372.1; -
 DR MGD; MGI:1891767; Chst7.
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 DR Transferrase
 KW SEQUENCE 484 AA: 54751 MW: 402C51ED185FDF8 CRC64;
 Query Match 30.6%; Score 624; DB 11; Length 484;
 Best Local Similarity 34.2%; Pred. No. 2.8e-50;
 Matches 136; Conservative 66; Mismatches 150; Indels 46; Gaps 7
 QY NISSLSMKAQPERMIVLVSSWRSGSSPFVGLFGCHPDVFLMEPAWHYMTFKOSTAM 89
 Db NLSAVGEAVTQEKOHIVYHATWRTSSSEFGLFNHPDVFLEYEPWHLMOALYPGDAES 146
 QY LHMAYRDLIRAVFLCDMSVFDAVMPGPRQ-----SSLFQWENSRAALCSAPAC 139
 Db LOGALRDLRLSFLRCDFSVLRLYAQGDGGERAPDSANLTMTLFRMRTNKVYCSPLCP 206
 QY IIPDQ-----IIPRAHCILCSQDPEVEYKCRSTSHYLYKEVRFNIOSTLYPLIKDS 195
 Db AAPRADVGVGEDKACSESTCPSPVSLRALEACRKPVVIVKIDVRLDLGVLPILLRDPG 266
 QY LNLHIVHLVPRPRAVERSRERTKGLMTDS-----RTVM-----GQHE 233
 Db LNLKVVQLFPRDRAVHNSRLSRQGLRSTIOVLTFRQGDHFHRYLHLHGVDARPGQA 326
 QY QKLKREDDPY---VMQVICSQSEIYKTIQSLSPKALQERYLVAVEDIARAPAQTSR 289
 Db RALPSAPRADEFLLTSALFVIGCAWLRDLFTTGAPAWMLRRRLRLRYEDLVNQPOQLRR 386
 QY MTEPVGLEFLHLQWVHNITRGKMG-DHAFHTARDAALNVSQAMRSLPEYKVSRLQ 348
 Db LLRSGLTLTALAALDAFAFNRMGASVYAGDRPHLSARARAVAHAMRELTSEQVRYET 446
 QY ACGDAMNLTGYRHVRSQEOORNL-----LDLSTWTV 381
 Db ACAPAMRLIATPRSGDEDRKTVRGCEPLETAKAMVAY 484

[illegible]

QY 245 ----VMOVICOSOLEIYKTIQSLPKALOERYLLVREYEDLARAFAVACTSRMYEFVGLFELP 300
 Db 340 FLTGALVEICEMALRDLLEFARGAPAWLRRLRYLREYEDLVROPRADRLRLRRSGLRALA 399
 QY 301 HLOTWVHNITRGKMG-DHAFTNARDALNVSOAMRMSLEVEYSRLQKAGDAMNLLGY 359
 Db 400 ALDAFALNMTRGAAAGADRFPHLSARDAREAVHAMRELSREQVRQVEVACAPAMRLAY 459
 QY 360 RHVRSOE 367
 Db 460 --PRSGEE 465

RESULT 12

Q9NS84 PRELIMINARY; PRT; 486 AA.
 AC Q9NS84; 01-OCT-2000 (Tremblrel, 15, Created)
 DT 01-OCT-2000 (Tremblrel, 15, Last sequence update)
 DE 01-MAR-2003 (Tremblrel, 23, Last annotation update)
 GN Chondroitin 6-sulfotransferase-2.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20347214; PubMed=10781596;
 RA Kitagawa H., Fujita M., Ito N., Sugahara K.;
 RT "Molecular cloning and expression of a novel chondroitin 6-O-
 RT sulfotransferase."
 RL J. Biol. Chem. 275:21075-21080(2000).
 DR EMBL: AB037187; BAA03217.1; -
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 KW Sulfotransferase.
 SQ SEQUENCE 486 AA; 54294 MW; B524D1430B3C9075 CRC64;

Query Match

Best Local Similarity 29.7%; Score 605.5; DB 4; Length 486;
 Best Local Similarity 37.0%; Pred. No. 1.6e-48;
 Matches 136; Conservative 56; Mismatches 133; Indels 43; Gaps 8;

QY 41 ERNHVYLSSMRSGSFFVGLFGQHPDVYFLMEPAMVMTFKOSTAMLMHMAVRDLIRA 100
 Db 100 EKHIYVHNWKRGSSFFGLFNGHPDVFLYEPMHLMQALYPGDAESLQGLRDLKRS 139
 QY 101 VFLCDMSVFDAYMEPG-PRRQ-----SLFQWENSRAICAPACDIIPDE---I 146
 Db 160 LFRCDPSVLYLAPPGPARAPDPTANLTALFRWNTKVICSPPLCPAPARAEVGL 219
 QY 147 IPANRRLCSQPFVEVEKACRSYSHVLEKVERFNQSLYPLKDPSLNHLIHLVRD 206
 Db 220 VEDTACERSCPVAIRALEACRKYVYVYIKDVLIDLGLVPLLRPGINLRVQLFRD 279
 QY 207 PRAVFSRERTKGLMIDS-----RIYV-----GHEKRLKEQPPY 244
 Db 280 PRAVHNSRLKSRGLRESIQVLRTRQGRDFRHLVLAHGVARPGQSALPAAPADP 339
 QY 245 ----VMOVICOSOLEIYKTIQSLPKALOERYLLVREYEDLARAFAVACTSRMYEFVGLFELP 300
 Db 340 FLTGALVEICEMALRDLLEFARGAPAWLRRLRYLREYEDLVROPRADRLRLRRSGLRALA 399
 QY 301 HLOTWVHNITRGKMG-DHAFTNARDALNVSOAMRMSLEVEYSRLQKAGDAMNLLGY 359
 Db 400 ALDAFALNMTRGAAAGADRFPHLSARDAREAVHAMRELSREQVRQVEVACAPAMRLAY 459
 QY 360 RHVRSOE 367
 Db 460 --PRSGEE 465

RESULT 13

Q98276 PRELIMINARY; PRT; 530 AA.
 AC Q98276; 01-NOV-1998 (Tremblrel, 08, Created)
 DT 01-NOV-1998 (Tremblrel, 08, Last sequence update)
 DE 01-DEC-2001 (Tremblrel, 19, Last annotation update)
 GN N-acetylglucosamine-6-O-sulfotransferase long form.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=98380482; PubMed=9712885;
 RA Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
 RA Mitsuoka C., Kannagi R., Habuchi O., Muramatsu T.;
 RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
 RT sulfotransferase."
 RL J. Biol. Chem. 273:22577-22583(1998).
 DR EMBL: AB011452; BAA32138.1; -
 DR EMBL: AB011452; BAA32139.1; -
 DR EMBL: AB011451; BAA32137.1; -
 DR MGI:1891160; Chst2.
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 KW Sulfotransferase.
 SQ SEQUENCE 530 AA; 57814 MW; A113E1B735C63BC CRC64;

Query Match

Best Local Similarity 29.4%; Score 598.5; DB 11; Length 530;
 Best Local Similarity 36.8%; Pred. No. 8e-48;
 Matches 134; Conservative 64; Mismatches 137; Indels 29; Gaps 6;

QY 41 ERNHVYLSSMRSGSFFVGLFGQHPDVYFLMEPAMVMTFKOSTAMLMHMAVRDLIRA 100
 Db 163 KROLVYFTWTRSGSFFGLFNGHPDVFLYEPVHVMQKLYPGDAVSLQGAARDMLSA 222
 QY 101 VFLCDMSVFDAYMEPGRRS-----SLFQWENSRAICAPACDIIPDEIIPAHCHLL- 155
 Db 223 LYRCDLSVFLYSPAGSGGRNLTTLTGIFGATKNCVSSPLCPAY-RKEVGLVDDRVCK 281
 QY 156 -CSQPFVEVEKACRSYSHVLEKVERFNQSLYPLKDPSLNHLIHLVRDPAVRSR 214
 Db 282 KCPQRRLAREEBCRKRTYVYIKGVRFDAVLAFLKDPALDKVHLVDRPAVASSR 341
 QY 215 ERTKGLMIDSRIVM-----GHEKRLKEQDQ---PY---VMOVICOS 252
 Db 342 IRRHGLIRSLQVRSRDRPRAHMPLEAAGHRLGAKKRGKMGADYHALGMEVYCN 401
 QY 253 QLEIYKTIQSLPKALOERYLLVREYEDLARAFAVACTSRMYEFVGLFELPRLQTVHNITRG 312
 Db 402 MARTLQALOPPMLOGHVLYVREYEDLVGPVTLRKYVDFVGLVSPENQOPALNMTSG 461
 QY 313 KGKGDHAFHTNARDALNVSOAMRMSLEVEYSRLQKAGDAMNLLGYRHVRSOEORNL 372
 Db 462 SGSSKPFVVSARNATQANAMRTALTFQDIKQVEEFCYPMALVGLTERVNSPEVEKDL 521
 QY 373 LDLL 376
 Db 522 KTL 525

RESULT 14

Q9UED5 PRELIMINARY; PRT; 483 AA.
 ID Q9UED5
 AC Q9UED5; 01-MAY-2000 (Tremblrel, 13, Created)
 DT 01-MAY-2000 (Tremblrel, 13, Last sequence update)
 DE 01-MAR-2003 (Tremblrel, 23, Last annotation update)
 GN N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST).
 GN G6ST.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:41:19 ; Search time 16.4574 Seconds
(without alignments) 992.382 Million cell updates/sec

Title: US-09-645-078-2

Perfect score: 2038

Sequence: 1 MLPRKKMLLFLVSQMAIL.....EQRNLLDLSTWTPVQIH 386

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 4231058 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Issued_Patents_AA:*
- 2: /cgn2_6/ptodata/1/1aa/5A-COMB.pdp:*
- 3: /cgn2_6/ptodata/1/1aa/5B-COMB.pdp:*
- 4: /cgn2_6/ptodata/1/1aa/6A-COMB.pdp:*
- 5: /cgn2_6/ptodata/1/1aa/PCTUS-COMB.pdp:*
- 6: /cgn2_6/ptodata/1/1aa/Backfile1.pdp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2038	100.0	386	3	US-09-045-284A-2
2	2038	100.0	386	4	US-09-190-911-1
3	1936	95.0	386	4	US-09-786-240-11
4	598.5	29.4	483	3	US-09-263-023-2
5	598.5	29.4	483	3	US-09-471-867-2
6	587.5	28.8	484	3	US-09-263-023-4
7	587.5	28.8	484	4	US-09-471-867-4
8	549	26.9	411	4	US-09-015-188-2
9	500.5	24.6	438	2	US-08-655-878-2
10	482	23.7	479	2	US-08-899-514-2
11	100.5	4.9	668	4	US-09-811-469-6
12	94.5	4.6	848	4	US-09-540-824-27
13	92	4.5	359	3	US-09-150-133-11
14	92	4.5	359	3	US-09-150-133-11
15	92	4.5	359	3	US-09-374-493-11
16	92	4.5	359	3	US-09-374-824-11
17	92	4.5	359	3	US-09-374-492-11
18	87.5	4.3	380	3	US-09-150-133-9
19	87.5	4.3	380	3	US-09-150-141-9
20	87.5	4.3	380	3	US-09-374-493-9
21	87.5	4.3	380	3	US-09-374-824-9
22	87.5	4.3	380	3	US-09-374-492-9
23	86.5	4.2	831	1	US-08-073-384C-5
24	86.5	4.2	831	1	US-08-254-359A-5
25	86.5	4.2	831	1	US-08-483-043-5
26	86.5	4.2	831	1	US-08-481-238-5
27	86.5	4.2	831	2	US-08-471-066B-5

28	86.5	4.2	831	2	US-08-484-956-5	Sequence 5, Appl1
29	86.5	4.2	831	2	US-08-757-653-5	Sequence 5, Appl1
30	86.5	4.2	831	2	US-08-599-491-5	Sequence 5, Appl1
31	86.5	4.2	831	2	US-08-756-386-5	Sequence 5, Appl1
32	86.5	4.2	831	2	US-08-823-516-5	Sequence 5, Appl1
33	86.5	4.2	831	3	US-08-682-853A-5	Sequence 5, Appl1
34	86.5	4.2	831	3	US-08-759-038-5	Sequence 5, Appl1
35	86.5	4.2	831	3	US-08-758-314-5	Sequence 5, Appl1
36	86.5	4.2	831	4	US-09-350-309-5	Sequence 5, Appl1
37	86.5	4.2	831	4	US-08-520-946-5	Sequence 5, Appl1
38	86.5	4.2	831	4	US-09-684-938-5	Sequence 5, Appl1
39	86.5	4.2	831	4	US-09-308-825A-1	Sequence 5, Appl1
40	84.5	4.1	2549	4	US-08-265-967C-1	Sequence 1, Appl1
41	84.5	4.1	2549	4	US-08-305-790B-2	Sequence 2, Appl1
42	83	4.1	195	4	US-09-252-991A-30847	Sequence 30847, A
43	83	4.1	222	2	US-08-485-721-9	Sequence 9, Appl1
44	83	4.1	222	2	US-08-392-935-9	Sequence 9, Appl1
45	83	4.1	222	5	PCT-US93-08326-9	Sequence 9, Appl1

ALIGNMENTS

RESULT 1									
US-09-045-284A-2									
: Sequence 2, Application US/09045284A									
: Patent No. 6265192									
: GENERAL INFORMATION:									
: APPLICANT: Bistrup, Annette									
: APPLICANT: Rosen, Steven D.									
: TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3									
: FILE REFERENCE: 6510-107US1									
: CURRENT APPLICATION NUMBER: US/09/045,284A									
: CURRENT FILING DATE: 1998-03-20									
: NUMBER OF SEQ ID NOS: 9									
: SOFTWARE: FastSeq for Windows Version 4.0									
: SEQ ID NO 2									
: LENGTH: 386									
: TYPE: PRT									
: ORGANISM: Homo sapiens									
US-09-045-284A-2									
Query Match 100.0%; Score 2038; DB 3; Length 386;									
Best Local Similarity 100.0%; Pred. No. 2,3e-212;									
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MLPRKKMLLFLVSQMAILALFFHMTSHNITSLSMKRAOPERMHVLVLSRRSGSFTVQ	60						
DB	1	MLPRKKMLLFLVSQMAILALFFHMTSHNITSLSMKRAOPERMHVLVLSRRSGSFTVQ	60						
QY	61	LEGGHPVEYLMERPAHVMVFQSTAMLMHAYRDLIRAVFLCDMSVFPAVMRPGRRQ	120						
DB	61	LEGGHPVEYLMERPAHVMVFQSTAMLMHAYRDLIRAVFLCDMSVFPAVMRPGRRQ	120						
QY	121	SSLQWENSRLACAPACDIIPODEIIPRACHRLCSQOPEVEVEKACRSYSHVLEVR	180						
DB	121	SSLQWENSRLACAPACDIIPODEIIPRACHRLCSQOPEVEVEKACRSYSHVLEVR	180						
QY	181	FFNIQSLYPLKDPSLNLHIVLVDRPAVFRSEERKGDIMDSRIYMGHEOKKED	240						
DB	181	FFNIQSLYPLKDPSLNLHIVLVDRPAVFRSEERKGDIMDSRIYMGHEOKKED	240						
QY	241	OPYVMOVITCOSOLEIKTIOSLPKALOERLLVRYEDLARAFAVOTSRMYEFGLEFLP	300						
DB	241	OPYVMOVITCOSOLEIKTIOSLPKALOERLLVRYEDLARAFAVOTSRMYEFGLEFLP	300						
QY	301	HLQTVWNITRGKMGDHAFTNARDALNVSQAMRWSLPEKYSRLQKAGDANMLLGYR	360						
DB	301	HLQTVWNITRGKMGDHAFTNARDALNVSQAMRWSLPEKYSRLQKAGDANMLLGYR	360						
QY	361	HVRSQEQORNLLDLSTWTPVQIH 386							
DB	361	HVRSQEQORNLLDLSTWTPVQIH 386							

Db 361 HVRSDEQRNLLDLSTWTPVPEQIH 386

RESULT 2

US-09-190-911-1

Sequence 1, Application US/09190911

Patent No. 6353365

GENERAL INFORMATION:

APPLICANT: Bistrup, Annette

APPLICANT: Rosen, Steven D.

APPLICANT: Tangemann, Kirsten

APPLICANT: Hemmerich, Stefan

TITLE OF INVENTION: GLYCOSTL SULFOTRANSFERASE-3

FILE REFERENCE: 6510-107CIP

CURRENT APPLICATION NUMBER: US/09/190,911

CURRENT FILING DATE: 1998-11-12

EARLIER APPLICATION NUMBER: 09/045,284

EARLIER FILING DATE: 1998-03-20

NUMBER OF SEQ ID NOS: 8

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 1

LENGTH: 386

TYPE: PRT

ORGANISM: H. sapiens

US-09-190-911-1

Query Match 100.0%; Score 2038; DB 4; Length 386;

Best Local Similarity 100.0%; Pred. No. 2,3e-212;

Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLPPKMKLLFLVSOMATLALFPHYSHNISLSKKAQPERMHVLYSSWRSQSSFTVGQ 60

DB 1 MLPPKMKLLFLVSOMATLALFPHYSHNISLSKKAQPERMHVLYSSWRSQSSFTVGQ 60

QY 61 LFGQHPDVEYLMPEAHVMTFKOSTAWMLHNAVRLIRAVFLCDMSVFDAYMEGPRRQ 120

DB 61 LFGQHPDVEYLMPEAHVMTFKOSTAWMLHNAVRLIRAVFLCDMSVFDAYMEGPRRQ 120

QY 61 LFGQHPDVEYLMPEAHVMTFKOSTAWMLHNAVRLIRAVFLCDMSVFDAYMEGPRRQ 120

DB 61 LFGQHPDVEYLMPEAHVMTFKOSTAWMLHNAVRLIRAVFLCDMSVFDAYMEGPRRQ 120

QY 121 SSLFQWENSRALCSAPACDIIPODEIIPRAHRLCSQOPFEVEKACRSYSHVYLKEVR 180

DB 121 SSLFQWENSRALCSAPACDIIPODEIIPRAHRLCSQOPFEVEKACRSYSHVYLKEVR 180

QY 121 SSLFQWENSRALCSAPACDIIPODEIIPRAHRLCSQOPFEVEKACRSYSHVYLKEVR 180

DB 121 SSLFQWENSRALCSAPACDIIPODEIIPRAHRLCSQOPFEVEKACRSYSHVYLKEVR 180

QY 181 FENLOSIVPLKDPSSLNLHIVLVDRPRAVFRSRETKDGLMIDSRIVNGQHEOKLKED 240

DB 181 FENLOSIVPLKDPSSLNLHIVLVDRPRAVFRSRETKDGLMIDSRIVNGQHEOKLKED 240

QY 181 FENLOSIVPLKDPSSLNLHIVLVDRPRAVFRSRETKDGLMIDSRIVNGQHEOKLKED 240

DB 181 FENLOSIVPLKDPSSLNLHIVLVDRPRAVFRSRETKDGLMIDSRIVNGQHEOKLKED 240

QY 241 QPYVWQVICSQOLEIKYTIQSLPKALOERYLLVREYEDLARAPVAQTSMEYFVGLFPL 300

DB 241 QPYVWQVICSQOLEIKYTIQSLPKALOERYLLVREYEDLARAPVAQTSMEYFVGLFPL 300

QY 241 QPYVWQVICSQOLEIKYTIQSLPKALOERYLLVREYEDLARAPVAQTSMEYFVGLFPL 300

DB 241 QPYVWQVICSQOLEIKYTIQSLPKALOERYLLVREYEDLARAPVAQTSMEYFVGLFPL 300

QY 301 HIQTWVHNTTRCKGMGDHAFHTNARDALNVSAWMSLPEYKVSRLQKACGDAMNLLGYR 360

DB 301 HIQTWVHNTTRCKGMGDHAFHTNARDALNVSAWMSLPEYKVSRLQKACGDAMNLLGYR 360

QY 301 HIQTWVHNTTRCKGMGDHAFHTNARDALNVSAWMSLPEYKVSRLQKACGDAMNLLGYR 360

DB 301 HIQTWVHNTTRCKGMGDHAFHTNARDALNVSAWMSLPEYKVSRLQKACGDAMNLLGYR 360

QY 361 HVRSDEQRNLLDLSTWTPVPEQIH 386

DB 361 HVRSDEQRNLLDLSTWTPVPEQIH 386

RESULT 3

US-09-786-240-11

Sequence 11, Application US/09786240

Patent No. 6558935

GENERAL INFORMATION:

APPLICANT: INCYTE PHARMACEUTICALS, INC.

APPLICANT: TANG, Y. TOM

APPLICANT: CORLEY, Neil C.

APPLICANT: GUEGLER, Karl J.

APPLICANT: BAUGHN, Mariah R.

APPLICANT: LAL, Preeti

APPLICANT: YUE, Henry

APPLICANT: HILLMAN, Jennifer L.

APPLICANT: AZIMZAI, Yalda

TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS

FILE REFERENCE: PF-0592 PCT

CURRENT APPLICATION NUMBER: US/09/786,240

CURRENT FILING DATE: 2002-03-12

PRIOR APPLICATION NUMBER: 09/150,657; unaassigned: 09/186,779; unaassigned: 60/133,6

PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1998-11-04; 1999-05-11

NUMBER OF SEQ ID NOS: 33

SOFTWARE: PERL Program

SEQ ID NO 11

LENGTH: 386

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

OTHER INFORMATION: Incyte ID No. 6558935 2617407CD1

US-09-786-240-11

Query Match 95.0%; Score 1936; DB 4; Length 386;

Best Local Similarity 95.6%; Pred. No. 2,6e-201;

Matches 369; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

QY 1 MLPPKMKLLFLVSOMATLALFPHYSHNISLSKKAQPERMHVLYSSWRSQSSFTVGQ 60

DB 1 MLPPKMKLLFLVSOMATLALFPHYSHNISLSKKAQPERMHVLYSSWRSQSSFTVGQ 60

QY 61 LFGQHPDVEYLMPEAHVMTFKOSTAWMLHNAVRLIRAVFLCDMSVFDAYMEGPRRQ 120

DB 61 LFGQHPDVEYLMPEAHVMTFKOSTAWMLHNAVRLIRAVFLCDMSVFDAYMEGPRRQ 120

QY 61 LFGQHPDVEYLMPEAHVMTFKOSTAWMLHNAVRLIRAVFLCDMSVFDAYMEGPRRQ 120

DB 61 LFGQHPDVEYLMPEAHVMTFKOSTAWMLHNAVRLIRAVFLCDMSVFDAYMEGPRRQ 120

QY 121 SSLFQWENSRALCSAPACDIIPODEIIPRAHRLCSQOPFEVEKACRSYSHVYLKEVR 180

DB 121 SSLFQWENSRALCSAPACDIIPODEIIPRAHRLCSQOPFEVEKACRSYSHVYLKEVR 180

QY 121 SSLFQWENSRALCSAPACDIIPODEIIPRAHRLCSQOPFEVEKACRSYSHVYLKEVR 180

DB 121 SSLFQWENSRALCSAPACDIIPODEIIPRAHRLCSQOPFEVEKACRSYSHVYLKEVR 180

QY 181 FENLOSIVPLKDPSSLNLHIVLVDRPRAVFRSRETKDGLMIDSRIVNGQHEOKLKED 240

DB 181 FENLOSIVPLKDPSSLNLHIVLVDRPRAVFRSRETKDGLMIDSRIVNGQHEOKLKED 240

QY 181 FENLOSIVPLKDPSSLNLHIVLVDRPRAVFRSRETKDGLMIDSRIVNGQHEOKLKED 240

DB 181 FENLOSIVPLKDPSSLNLHIVLVDRPRAVFRSRETKDGLMIDSRIVNGQHEOKLKED 240

QY 241 QPYVWQVICSQOLEIKYTIQSLPKALOERYLLVREYEDLARAPVAQTSMEYFVGLFPL 300

DB 241 QPYVWQVICSQOLEIKYTIQSLPKALOERYLLVREYEDLARAPVAQTSMEYFVGLFPL 300

QY 241 QPYVWQVICSQOLEIKYTIQSLPKALOERYLLVREYEDLARAPVAQTSMEYFVGLFPL 300

DB 241 QPYVWQVICSQOLEIKYTIQSLPKALOERYLLVREYEDLARAPVAQTSMEYFVGLFPL 300

QY 301 HIQTWVHNTTRCKGMGDHAFHTNARDALNVSAWMSLPEYKVSRLQKACGDAMNLLGYR 360

DB 301 HIQTWVHNTTRCKGMGDHAFHTNARDALNVSAWMSLPEYKVSRLQKACGDAMNLLGYR 360

QY 301 HIQTWVHNTTRCKGMGDHAFHTNARDALNVSAWMSLPEYKVSRLQKACGDAMNLLGYR 360

DB 301 HIQTWVHNTTRCKGMGDHAFHTNARDALNVSAWMSLPEYKVSRLQKACGDAMNLLGYR 360

QY 361 HVRSDEQRNLLDLSTWTPVPEQIH 386

DB 361 HVRSDEQRNLLDLSTWTPVPEQIH 386

RESULT 4

US-09-263-023-2

Sequence 2, Application US/09263023

Patent No. 6037159

GENERAL INFORMATION:

APPLICANT: Uchimura, Kenji

APPLICANT: Muramatsu, Hideki

APPLICANT: Kadomatsu, Kenji

APPLICANT: Kannagi, Reiichi

APPLICANT: Habuchi, Osami

APPLICANT: Muramatsu, Takashi

TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND

FILE REFERENCE: TOYAMA1.001AUS

CURRENT APPLICATION NUMBER: US/09/263,023

CURRENT FILING DATE: 1999-03-05

EARLIER APPLICATION NUMBER: JP 10-54007

EARLIER FILING DATE: 1998-03-05

EARLIER APPLICATION NUMBER: JP 10-177844

EARLIER FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 10

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 2


```

; LENGTH: 483
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-263-023-2

```

Query Match	29.48;	Score 598.5;	DB 3;	Length 483;
Best Local Similarity	36.88;	Pred. No. 3.6e-56;		
Matches 134;	Conservative 64;	Mismatches 137;	Indels 29;	Gaps 6

[illegible]

```

RESULT 5
US-09-471-867-2
; Sequence 2, Application US/09471867
; Patent No. 6455289
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kanagaki, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; TITLE OF INVENTION: DNA ENCODING THE SAME
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/471.867
; CURRENT FILING DATE: 1999-12-23
; EARLIER APPLICATION NUMBER: US 09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-471-867-2

```

Query Match	29.4%;	Score 598.5;	DB 4;	Length 483;
Best Local Similarity	36.8%;	Pred. No. 3.6e-56;		
Matches 134;	Conservative 64;	Mismatches 137;	Indels 25;	Gaps 6
07	41	ERMHYLVLSMRSGSFVQLEGGQHPDVFYILMEPMHNYMTFKQSTAMMLHNAVRLDITA	100	

[illegible]

```

: RESULT 6
: US-09-263-023-4
: Sequence 4, Application US/09263023
: Patent No. 6037159
: GENERAL INFORMATION:
: APPLICANT: Uchimura, Kenji
: APPLICANT: Muramatsu, Hideki
: APPLICANT: Kadomatsu, Kenji
: APPLICANT: Kannagi, Reiji
: APPLICANT: Habuchi, Osami
: APPLICANT: Muramatsu, Takashi
: TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGALACTOSAMINE-6-0-SULFOTRANSFERASE AND
: TITLE OF INVENTION: DNA ENCODING THE SAME
: FILE REFERENCE: TOYAMA1.001AUS
: CURRENT APPLICATION NUMBER: US/09/263.023
: CURRENT FILING DATE: 1998-03-05
: EARLIER APPLICATION NUMBER: JP 10-54007
: EARLIER FILING DATE: 1998-03-05
: EARLIER APPLICATION NUMBER: JP 10-177844
: EARLIER FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 4
: LENGTH: 484
: TYPE: prt
: ORGANISM: Homo sapiens
: US-09-263-023-4

```

Query Match	28.8%	Score 587.5	DB 3	Length 484
Best Local Similarity	36.0%	Pred. No. 5	6e-55	
Matches 131	Conservative	67	Mismatches 137	Indels 29
				Gaps 6
QY	41	ERMHVLYLSSWRSGSSFVGOLFQHPDVEFLMEPRAHVWMTPEKOSTAMWLAHVADLIRA	100	
		: : : : : : : : :		
DB	117	KRHMMVYPTTWRSGSFFGELFNQNPVEVFELTEPVYHWQKLYPDGDAVSLQAGARDMLISA	176	
		: : : : : : : : :		
QY	101	VFLCDMSYFDAYMEGRPRRS-----SLQWEMSRALCSAPACDIIPODEIIPARICRL-	155	
		: : : : : : : : :		
DB	177	LYRCDLSEFOLYSPAGSGGRMLTTGLIGCAATNKKVYCSSPLCPAY-RKEWGLVDDRYCK	235	
		: : : : : : : : :		
QY	156	-CSQOPFEVVEKACSYSHVYLKEVRFNLQSLYLDQPSLNTHIVHLVYRDPAYFRSR	214	
		: : : : : : : : :		
DB	236	KCPQRLRAFEEBECKRYTLVYKGVYRDVAVLALDLNRPALDLKVIHLVYRDPAYASR	295	
		: : : : : : : : :		
QY	215	ERTKGDLMLDSRIYV-----GQHEQKLKKE-----DQPYV---VMQVICS	252	
		: : : : : : : : :		

```

Db      296  IRRRHGLIRESLQVYRSRDPRAHMPFLEAGHKIGAKKEGVCBPADYHALGAMETICNS 355
OY      253  QLEIKYKTSUPKALQERLYRYEDLARAQAQTSRMVEFGLEFLPHLOQWVNIITRG 312
Db      356  MAKTIQTLQPPDWLQGHILVRYEDLVGDPVKTLRRYIDFVGLVSPMEQFALNMTSG 415
OY      313  KMGDHAFTNARDALNVSQAWMSLPEKYVSRLOKACGDAMNLLGYRHVRSEQORNL 372
Db      416  SSSSSKPFVYSARNNTQAANMARTALTTFQIQKQVEEFCQPAVAVIGYEVNPSPEVKDLS 475
OY      373  LDLL 376
Db      476  KTL 479

```

RESULT 7

```

-09-471-867-4
Sequence 4, Application US/09471867
Patent No. 6455289
GENERAL INFORMATION:
APPLICANT: Uchimura, Kenji
APPLICANT: Muramatsu, Hideki
APPLICANT: Kadomatsu, Kenji
APPLICANT: Kamegaki, Reiji
APPLICANT: Habuchi, Osami
APPLICANT: Muramatsu, Takashi
TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
FILE REFERENCE: TOYAMA1.001AUS
CURRENT APPLICATION NUMBER: US/09/411,867
CURRENT FILING DATE: 1999-12-23
EARLIER APPLICATION NUMBER: US 09/263,023
EARLIER FILING DATE: 1999-03-05
EARLIER APPLICATION NUMBER: JP 10-54007
EARLIER FILING DATE: 1998-03-05
EARLIER APPLICATION NUMBER: JP 10-177844
EARLIER FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 484
TYPE: PRP
ORGANISM: Homo sapiens
US-09-471-867-4

```

Query Match 28.8%; Score 587.5; DB 4; Length 484;
 Best Local Similarity 36.0%; Pred. No. 5,6e-55;
 Matches 131; Conservative 67; Mismatches 137; Indels 29; Gaps 6;

```

OY      41  EEMHVLVLSWRSGSFGQLFGQHPDVFTLMEPAHVMYMTKOSTAMMLHNAVRLIRA 100
Db      117  KHNMYVFTTWRSFGSFGELFNONPEVFLPEYVHWQKLYPDGAVSLQGAARMLSA 176
OY      101  VFLCDMSYVDAYMEGPRROS----SLFQWENSRLCSAPACDIIPODEIIPRACHRL- 155
Db      177  LYRCLSLVQLYSPAGSGGRNLTIGIFGAATNKVCCSPCLCPAY-KREVLVLVDRCK 235
OY      156  -CSQGFEEVVERACRSYSHVYLKEVRFENLQSLYPLDLKPSLNHLVHVLVDRPRAVRSR 214
Db      236  KCPQRLARFEECECKRYFLVIGVRFVAVLAPLRLPDALDVYHLVVRPRAVASR 295
OY      215  ERTKDLMDISRYV-----GQHEOKLKE---DQPIY---VMQYICOS 252
Db      296  IRRRHGLIRESLQVYRSRDPRAHMPFLEAGHKIGAKKEGVCBPADYHALGAMETICNS 355
OY      253  QLEIKYKTSUPKALQERLYRYEDLARAQAQTSRMVEFGLEFLPHLOQWVNIITRG 312
Db      356  MAKTIQTLQPPDWLQGHILVRYEDLVGDPVKTLRRYIDFVGLVSPMEQFALNMTSG 415
OY      313  KMGDHAFTNARDALNVSQAWMSLPEKYVSRLOKACGDAMNLLGYRHVRSEQORNL 372
Db      416  SSSSSKPFVYSARNNTQAANMARTALTTFQIQKQVEEFCQPAVAVIGYEVNPSPEVKDLS 475

```

```

OY      373  LDLL 376
Db      476  KTL 479

```

RESULT 8

```

US-09-015-188-2
Sequence 2, Application US/09015188C
Patent No. 6393358
GENERAL INFORMATION:
APPLICANT: Williams, Kevin J
APPLICANT: Tabas, Ira
TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
FILE REFERENCE: JEPF-0231
CURRENT APPLICATION NUMBER: US/09/015,188C
CURRENT FILING DATE: 1998-01-29
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 411
TYPE: PRP
ORGANISM: Homo sapiens
US-09-015-188-2

```

Query Match 26.9%; Score 549; DB 4; Length 411;
 Best Local Similarity 34.3%; Pred. No. 6,5e-51;
 Matches 139; Conservative 64; Mismatches 160; Indels 42; Gaps 10;

```

OY      6  KKLLLFVLS---QNALILPFRMTSHNISLSMKAQPERM-----HYLV 47
Db      6  KAVLLILALASIAIQYTAIRFTYAKSFHCPGLAEGALRCEESPTAYNLRSKTHILI 65
OY      48  ISSWRSGSFGQLFGQHPDVFTLMEPAHVMYMTKOSTAMMLHNAVRLIRA 99
Db      66  LATTSSGSSFGQLNGHLDVFTLEPRLHYQNTLIPFTQKSRADRYMLGASRDLLR 125
OY      100  AVFLCDMSYVDAYMEGPRROS--LFQWENSRLCSAPACDIIPODEIIPRACHRLC 156
Db      126  SLYDCDLVFLFENYIKRPVNHHTDRIFRGASRVLCSPVCDPPGADVLVEEGCVKRC 185
OY      157  SQGFEEVVERACRSYSHVYLKEVRFENLQSLYPLDLKPSLNHLVHVLVDRPRAVRSR 216
Db      186  GLNLTVAEACRERSHAIKTVRVENVDLRALVEDPRLNKVTLQVLRDPGGLIASNSE 245
OY      217  TKGDLMDISRYVGOHEOKLKEKEDQPYV---MOVICOSLEIKYKTSUPKALQERYL 272
Db      246  TFRDYYRLRLMYGGR-----KPYNLDVQLTTCEDDFNSVSTGLMRPMLKGYM 298
OY      273  LVRYEDLARAQAQTSRMVEFGLEFLPHLOQWVNIITRG-KMGDHAFTNARDALNVS 331
Db      299  LVRYEDLARAQAQTSRMVEFGLEFLPHLOQWVNIITRG-KMGDHAFTNARDALNVS 357
OY      333  QAWMSLPEKYVSRLOKACGDAMNLLGYRHVRSEQORNL 376
Db      358  EKMRFRLSTDYVAFQANMCQOVLADQGYKIAASEELNPSVSLV 402

```

RESULT 9

```

US-08-655-878-2
Sequence 2, Application US/08655878
Patent No. 5827713
GENERAL INFORMATION:
APPLICANT: FUKUTA, MASAKAZU
APPLICANT: HABUCHI, OSAMI
TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESS:
STREET:
CITY:
STATE:

```

```

: COUNTRY:
: ZIP:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/655,878
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME:
: REGISTRATION NUMBER:
: REFERENCE/DOCKET NUMBER:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 458
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-655-878-2

```

Query Match 24.6%; Score 500.5; DB 2; Length 458;

Best Local Similarity 33.2%; Pred. No. 1.4e-45;

Matches 120; Conservative 75; Mismatches 133; Indels 33; Gaps 13;

```

QY 34 LSKADPERHNVIVLSSRGSSFFVGLFGQHPDVFLMPAHV--WMTFKOSTAMLH 91
DB 104 LGIAAPRRHVVLMATRTGSSFFVGFNOGNIFFLEPLMIERTVFEPEGANAVG 163
QY 92 MAV--RDLIRAVFLCDMSVFDAYMEPRRQ--SLFQWENSRLCSAPACDIIPQ-DEI 146
DB 164 SALVYRNVLOQLLDCDIYIESFTSPAPBEHLRALFRSSSHSLCEPVC--TPSLKVV 221
QY 147 IPRAHCR-LICSOQPEVEVEKACRSYSHVLAKEVRFNLQSLYPLKPSLNLHIVLVR 205
DB 222 FEYHYCKHNRRCGLNITLAEACRRKQHMALKTIVRIQLPLEADPRDLRIQLVLR 281
QY 206 DPRAVFRSRRTGDLMDIRIVMGHEOKL-----KEQPIYVMQVICS-QLE 255
DB 282 DPRAVLVS-----MV--AFSGKYESWKMAAEGLAEPLQDE-VQRLRGNCESIRLS 330
QY 256 IYKTIOSLPKALQERYLLVREYEDLARAPVAQTSRMVEFVGLLEFLPHLQTVWHTTRGKM 315
DB 331 AELGLNR-PWMLGRVLYVEDYARAPLKALEMIFAGIHPPVQVEEIRANTQAP-Q 388
QY 316 GDHAFHTNADALNVSQAWMSLPEYKVSRLQKACGDAMNLGGRHVSQEQORNLIDL 375
DB 389 DSGIYSTOKNSSEQEFKMFSLPKLAQVVOACPEAMRLFGYKLASSAQELTNRLSL 448
QY 376 L 376
DB 449 L 449

```

RESULT 10

US-08-899-514-2

; Sequence 2, Application us/08899514

; Patent No. 5910581

; GENERAL INFORMATION:

; APPLICANT: HABUCHI, OSAMI

; APPLICANT: FUKUTA, MASAKAZU

; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN

; TITLE OF INVENTION: SUBOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING

; NUMBER OF SEQUENCES: 9

```

: CORRESPONDENCE ADDRESS:
: ADDRESSEE: KNOBE, MARTENS, OLSON & BEAR, LLP
: STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
: CITY: NEWPORT BEACH
: STATE: CALIFORNIA
: COUNTRY: US
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/899,514
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: DANIEL E. ALTMAN
: REGISTRATION NUMBER: 34,115
: REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 714 760 0404
: TELEFAX: 714 760 9502
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 479
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-899-514-2

```

Query Match 23.7%; Score 482; DB 2; Length 479;

Best Local Similarity 32.4%; Pred. No. 1.5e-43;

Matches 113; Conservative 71; Mismatches 141; Indels 24; Gaps 10;

```

QY 42 RMHVLVSSWRSSGFFVGLFGQHPDVFLMPAHVWMT--FKOSTAMLHMAV--RDL 97
DB 132 RRHVLMAATRTGSSFFVGFNOGNIFFLEPLMIERTVFEPEGANASALVYRDV 191
QY 98 IRAVFLCDMSVFDAYMEPRRQSSSLFQWE--NSRALCSAPACDIIPQDEIIPRAHCR-L 154
DB 192 LKQFLDCDLYLEHFTPLPEDLTQFMFRGSSRLCEDPVCYFPVK-KVEEYHCKNR 250
QY 155 LCSQOPEVEVEKACRSYSHVLAKEVRFNLQSLYPLKPSLNLHIVLVRDPRVRSR 214
DB 251 RCGLNVTLAEACRRKREHMLKAVRIQLPLEADPRDLRIQLVDRPRAVLASR 310
QY 215 -----ERTGDLMDIRIVMGHEOKLKEQDPYVMQVICS-QLEIYKTIOSLPKAL 267
DB 311 MVAEPAGYKTKWKWLDDEGQDLNEEVOR-----LRNCESIRLSAELGLNQ-PAWL 362
QY 268 QERYLVREYEDLARAPVAQTSRMVEFVGLLEFLPHLQTVWHTTRGKMGAHFTNARDA 327
DB 363 RGRHVLRYVEDYARAPLKALEMIFAGIHPPVQVEEIRANTQAP-Q 421
QY 328 LNVSQAWMSLPEYKVSRLQKACGDAMNLGGRHVSQEQORNLIDL 376
DB 422 SEQEFKWFSPMFKLAQVVOACPEAMRLFGYKLARDAALTNRSVSL 470

```

RESULT 11

US-09-811-469-6

; Sequence 6, Application us/09811469

; Patent No. 6551809

; GENERAL INFORMATION:

; APPLICANT: YAN, Chunhua et al.

; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,

; FILE REFERENCE: C1001171

;; CURRENT APPLICATION NUMBER: US/09/811,469
 ;; CURRENT FILING DATE: 2001-03-20
 ;; NUMBER OF SEQ ID NOS: 11
 ;; SOFTWARE: FastSeq for Windows Version 4.0
 ;; SEQ ID NO 6
 ;; LENGTH: 668
 ;; TYPE: prt
 ;; ORGANISM: Human
 US-09-811-469-6

Query Match 4.9%; Score 100.5; DB 4; Length 668;
 Best Local Similarity 22.7%; Pred. No. 0.056;
 Matches 57; Conservative 35; Mismatches 94; Indels 65; Gaps 12;

QY 156 CSQPEVEVERKASYSHYVLEKEVREFNLSLYPLKDPSLMLHIYLVDPRAVRSRE 215
 DB 356 CTKKAIETECEADONMNVLLLEE--NASDCCILSS-----LVQAMDDPHC-----RT 402
 216 RTKGDLMTDSRYMG-----QHOKIKKEDQPYVMVQVICSQSLKTIQSLPKAL 267
 DB 403 RIGFOSLIQKEMVWGCHFLDRCNHRLRNDKEEVPEFLFLDCC-----VMQLVHOPPAF 457
 QY 268 Q--ERYLVREYEDLARAPYAOTSRYMEFEVLEFLPHLQTVWNITERGK--GMDHAFHTN 323
 DB 458 EETETVLYLVSDSL-----YIPITSEFFFNSPHQKDTNMGREGQDPTQ 499
 QY 324 ABDALNVSQARMSLEPYE-KVSRLOKACGDAMNL-----GYRVRSEQEQRNLL 373
 DB 500 SK-PNLTLVWDMVSVOFEPRKQTLTK-----NPLYVERPKLDKGRKMRKFOROLSL 552
 QY 374 DLTSTWVPEQ 384
 DB 553 PLTOSKSSPKR 563

RESULT 12

US-09-540-824-27
 ; Sequence 27, Application US/09540824
 ; Patent No. 6383753
 ; GENERAL INFORMATION:
 ; APPLICANT: Thiele, Dennis
 ; APPLICANT: Liu, Phillip
 ; TITLE OF INVENTION: No. 6383753el Yeast and Mammalian Regulators of Cell Proliferat
 ; FILE REFERENCE: UM-04266
 ; CURRENT APPLICATION NUMBER: US/09/540,824
 ; CURRENT FILING DATE: 2000-03-31
 ; NUMBER OF SEQ ID NOS: 28
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 27
 ; LENGTH: 848
 ; TYPE: prt
 ; ORGANISM: Caenorhabditis elegans
 US-09-540-824-27

Query Match 4.6%; Score 94.5; DB 4; Length 848;
 Best Local Similarity 21.2%; Pred. No. 0.36;
 Matches 74; Conservative 46; Mismatches 130; Indels 99; Gaps 15;

QY 5 KMKLLPLVSOMATLALFPHMYSHNISLSMKAAQ---PERMHVVLVSSMRGSSSFVQOL 61
 DB 426 KRLKMLFL-----CLRWFDDMSKELTAEVYKKNLQETMHMLKPKDIOGVGRVRL 481
 QY 62 FGOH-----PDVEYLMPEAMHWMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFD 110
 DB 482 IQHKKRGODKQKSSPVSGFLISAIRLVSGLPFVADSV--HPVV--VPALFLATEALCS 536
 QY 111 A-----YMEGPRROSSLPQWENSRALCSAP-----ACDIIPD 144
 DB 537 AACANLMLAKQIQOLANALIVEYSESKRYPPELVAFARSALLIATYKSEKPAINGFP-- 594
 QY 145 EIIPIAHCRLLC-----SQQPF-----EVEKACRSYSHVLYKCV 179
 DB 595 --ISKPHMLCFEKKLFLITKNYMWISOKEFDDISSFNRYIDELYLKMTKRTTGPALOPT 652

QY 180 REFINQSLYPLKDPSLMLHIYH---LVDRPRAVRSRERTKGDLMID-SRIYMGQHEQ 234
 DB 653 SLTTFINNSP--SDPSKLHVRALLSTLQHLRVAYSQNETYYSIVKFPFLRLSISQAK 710
 QY 235 KKKKEDQPYVMVQVICS-----QLEIKTIQSLKALQERY 271
 DB 711 NLPAEVOE--ELETICASKMAEIGAKCRLVHLISLVKTEKSMKMLEPRP 757

RESULT 13

US-09-150-133-11
 ; Sequence 11, Application US/09150133B
 ; Patent No. 6060295
 ; GENERAL INFORMATION:
 ; APPLICANT: The Board of Regents of the University of Oklahoma
 ; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
 ; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
 ; FILE REFERENCE: 5820.504
 ; CURRENT APPLICATION NUMBER: US/09/150,133B
 ; CURRENT FILING DATE: 1998-09-09
 ; NUMBER OF SEQ ID NOS: 45
 ; SOFTWARE: Wordperfect 5.1 (saved in Ascii format)
 ; SEQ ID NO 11
 ; LENGTH: 359
 ; TYPE: prt
 ; ORGANISM: Caenorhabditis elegans
 US-09-150-133-11

Query Match 4.5%; Score 92; DB 3; Length 359;
 Best Local Similarity 19.5%; Pred. No. 0.18;
 Matches 63; Conservative 48; Mismatches 114; Indels 98; Gaps 15;

QY 52 RSGSSFVQGLFGQHPDV---FYLMPEAMHWMTFKOSTAMMLHMAVRDLIRAVFLCDMS 107
 DB 95 RSGTFLMRALIDHAPDVRCGETMLPISFLTWQAGWRND-WVNNNSGI-----TQE 143
 QY 108 VFDAVMEPPRROSSLPQWENSRALCSAPACDIIPQ-DEIIRAHCRLLCSQPEVYK 166
 DB 144 VFDDAV-----SAITTEIVAHSSLAER-----LCNDDP----- 172
 QY 167 ACSRSYHVLEKEVREFNLSLYPLKDPSLMLHIYLVDPRAVRSRERTK---GDLM 222
 DB 173-----YIALMLPILR-----RLYP-----NAKFIIMTDAVAHVSMIERKVPVAGYWT 216
 QY 223 IDSRIYMGQHEQKLEKEDQPYVMVQVICSQ---LEIKTIQSLKALQERYLLVRYED 278
 DB 217 SDEISMVQWNOELRK-----MFOCNNAPGQCICKY-----YER 251
 QY 279 LARAPYAOTSRYMEFEVLEFLPHLQTVWNITERGKMGHAFH-TNARDALNVSQARMS 337
 DB 252 LIQPAEELIRITNFDLPFSOOMLRHODLIDDEVLDNDEFSASQVKNISINTKALTSWF 311
 QY 338 LPY-EKVSRLQKACGDAMNLGY 359
 DB 312 DCFSEETLAKLDVAPFLGILGT 334

RESULT 14

US-09-150-141-11
 ; Sequence 11, Application US/09150141B
 ; Patent No. 6071732
 ; GENERAL INFORMATION:
 ; APPLICANT: The Board of Regents of the University of Oklahoma
 ; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
 ; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
 ; FILE REFERENCE: 5820.495
 ; CURRENT APPLICATION NUMBER: US/09/150,141B
 ; CURRENT FILING DATE: 1998-09-09
 ; NUMBER OF SEQ ID NOS: 45
 ; SOFTWARE: Wordperfect 5.1 (saved in Ascii format)
 ; SEQ ID NO 11
 ; LENGTH: 359

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:42:44 ; Search time 20.447 Seconds
(Without alignments)
2241.953 Million cell updates/sec

Title: US-09-645-078-2
Perfect score: 2038
Sequence: 1 MLPRKKKLLFLVSOMAIL.....EQRLLDLSTWPEQIH 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCY_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/PCYUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubppaa/US10C_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2038	100.0	386	9 US-09-816-825-2	Sequence 2, Appl1
2	2038	100.0	386	14 US-10-007-262-1	Sequence 5, Appl1
3	1028	50.4	418	9 US-09-927-602-5	Sequence 1, Appl1
4	1019	50.0	395	9 US-09-927-602-2	Sequence 2, Appl1
5	1013.5	49.7	395	9 US-09-927-602-3	Sequence 3, Appl1
6	1008	49.5	390	9 US-09-927-602-4	Sequence 4, Appl1
7	821	40.3	171	9 US-09-927-602-8	Sequence 8, Appl1
8	598.5	29.4	483	15 US-10-212-933-2	Sequence 4, Appl1
9	587.5	28.8	484	15 US-10-212-933-4	Sequence 25, App
10	519.5	25.5	531	9 US-09-833-790-255	Sequence 6, Appl1
11	519.5	25.5	169	9 US-09-927-602-6	Sequence 7, Appl1
12	511.5	25.1	169	9 US-09-927-602-7	Sequence 9, Appl1
13	336	16.5	169	9 US-09-927-602-9	Sequence 10, Appl1
14	335.5	16.5	179	9 US-09-927-602-10	Sequence 11, Appl1
15	293	14.4	174	9 US-09-927-602-11	Sequence 11, Appl1

16	116	5.7	387	15 US-10-126-279-21	Sequence 21, Appl1
17	100.5	4.9	668	11 US-09-811-469-6	Sequence 6, Appl1
18	90	4.4	393	15 US-10-126-279-2	Sequence 20, Appl1
19	89	4.4	320	9 US-09-854-122-20	Sequence 15, Appl1
20	89	4.4	1968	15 US-10-223-070-15	Sequence 1903, Ap
21	88	4.3	2209	10 US-09-902-941-1903	Sequence 1903, Ap
22	88	4.3	2209	10 US-09-849-626-1903	Sequence 1903, Ap
23	88	4.3	2209	15 US-10-017-754-1903	Sequence 5, Appl1
24	86.5	4.2	831	11 US-09-940-244-5	Sequence 5, Appl1
25	86.5	4.2	831	11 US-09-940-244-5	Sequence 5, Appl1
26	86.5	4.2	831	11 US-09-982-667-5	Sequence 5, Appl1
27	86.5	4.2	831	11 US-09-864-636A-158	Sequence 158, App
28	86.5	4.2	831	11 US-09-941-193A-5	Sequence 5, Appl1
29	86.5	4.2	831	14 US-10-033-297-5	Sequence 5, Appl1
30	86.5	4.2	831	14 US-10-081-806-5	Sequence 5, Appl1
31	86.5	4.2	831	15 US-10-074-228-5	Sequence 5, Appl1
32	86	4.2	824	10 US-09-801-368-312	Sequence 312, App
33	82.5	4.0	388	15 US-10-126-279-6	Sequence 6, Appl1
34	82.5	4.0	777	11 US-09-811-469-2	Sequence 2, Appl1
35	82.5	4.0	1140	11 US-09-950-634-4	Sequence 4, Appl1
36	82.5	4.0	2549	11 US-09-950-634-3	Sequence 3, Appl1
37	81.5	4.0	1463	15 US-10-226-065-3	Sequence 3, Appl1
38	81	4.0	228	11 US-09-813-153-286	Sequence 286, App
39	81	4.0	295	10 US-09-965-529-31	Sequence 31, Appl1
40	81	4.0	295	11 US-09-969-680A-31	Sequence 31, Appl1
41	81	4.0	1100	9 US-09-815-242-11861	Sequence 53, Appl1
42	79.5	3.9	177	8 US-08-424-5508-53	Sequence 574, App
43	79.5	3.9	840	11 US-09-864-636A-574	Sequence 3, Appl1
44	79.5	3.9	1458	15 US-10-203-224-3	Sequence 11350, A
45	79	3.9	402	9 US-09-815-242-11350	

ALIGNMENTS

RESULT 1
US-09-816-825-2
; Sequence 2, Application US/09816825
; Patent No. US20010051370A1
GENERAL INFORMATION:
APPLICANT: Bistrup, Annette
APPLICANT: Rosen, Steven D.
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
FILE REFERENCE: 6510-107CON
CURRENT APPLICATION NUMBER: US/09/816, 825
CURRENT FILING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: 09/045, 284
PRIOR FILING DATE: 1998-03-20
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 386
TYPE: PRT
ORGANISM: Homo sapiens
US-09-816-825-2

Query Match 100.0%; Score 2038; DB 9; Length 386;
Best Local Similarity 100.0%; Pred. No. 1.9e-202;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLPRKKKLLFLVSOMAILLFFPMYSHNTSSSKMAQPRMHLVLSWSSSPVQ 60
Db 1 MLPRKKKLLFLVSOMAILLFFPMYSHNTSSSKMAQPRMHLVLSWSSSPVQ 60
QY 61 LFGQHPDVEFYLMEDAMHVMTEFKOSTAMMLMAVRDLIRAVFLCDMSVFDAYPEGRRQ 120
Db 61 LFGQHPDVEFYLMEDAMHVMTEFKOSTAMMLMAVRDLIRAVFLCDMSVFDAYPEGRRQ 120
QY 121 SSLFQWENSRLCLCAPACDITIPDEITIPRAHCRLLCSQOPREVEYKACRSYHVLKEVR 180
Db 121 SSLFQWENSRLCLCAPACDITIPDEITIPRAHCRLLCSQOPREVEYKACRSYHVLKEVR 180

```

QY 181 FENLQSLVPLKDPISLNIHVLVRDPAVFRSRETKGDLMDISRTVWGQHOEQLKKED 240
DB 181 FENLQSLVPLKDPISLNIHVLVRDPAVFRSRETKGDLMDISRTVWGQHOEQLKKED 240
QY 241 QPYVWQVQCSOLEYTKTQISLPKALQRYLLVREDLARAFAVQTSRYEFGLEFLP 300
DB 241 QPYVWQVQCSOLEYTKTQISLPKALQRYLLVREDLARAFAVQTSRYEFGLEFLP 300
QY 301 HLOTWVHNITRGKMGDAFHTNARDALNVSOAMWSLPEYKVSRLQKACGDAMNLGGR 360
DB 301 HLOTWVHNITRGKMGDAFHTNARDALNVSOAMWSLPEYKVSRLQKACGDAMNLGGR 360
QY 361 HVRSEQORNLIDLSTVTPEQIH 386
DB 361 HVRSEQORNLIDLSTVTPEQIH 386

```

RESULT 2

```

S-10-007-262-1
Sequence 1, Application US/10007262
Publication No. US20020164748A1
GENERAL INFORMATION:
APPLICANT: Bistrup, Annette
APPLICANT: Rosen, Steven D.
APPLICANT: Tangemann, Kirsten
APPLICANT: Hemmerich, Stefan
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
FILE REFERENCE: 6510-107CIP
CURRENT APPLICATION NUMBER: US/10/007,262
CURRENT FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 386
TYPE: PRT
ORGANISM: H. sapiens
US-10-007-262-1

```

```

Query Match 100.0%; Score 2038; DB 14; Length 386;
Best Local Similarity 100.0%; Pred. No. 1.9e-202;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MLPPKMKLLFLVSMALILAFPHMYSHNISLSKKAQPERMHVLYLSSWRSGSFFVQ 60
b 1 MLPPKMKLLFLVSMALILAFPHMYSHNISLSKKAQPERMHVLYLSSWRSGSFFVQ 60
QY 61 LFGQHPDVYELMEPAHVMWTFKOSTAMMLHMAVRODLIRAVFLCDMSVDFDAYEPGPRQ 120
DB 61 LFGQHPDVYELMEPAHVMWTFKOSTAMMLHMAVRODLIRAVFLCDMSVDFDAYEPGPRQ 120
QY 121 SSLFQWENSRLCASAPACDIIPODEIIPRAHCRLLCSQPFVEVEKACRSYSHVLEKVR 180
DB 121 SSLFQWENSRLCASAPACDIIPODEIIPRAHCRLLCSQPFVEVEKACRSYSHVLEKVR 180
QY 181 FENLQSLVPLKDPISLNIHVLVRDPAVFRSRETKGDLMDISRTVWGQHOEQLKKED 240
DB 181 FENLQSLVPLKDPISLNIHVLVRDPAVFRSRETKGDLMDISRTVWGQHOEQLKKED 240
QY 241 QPYVWQVQCSOLEYTKTQISLPKALQRYLLVREDLARAFAVQTSRYEFGLEFLP 300
DB 241 QPYVWQVQCSOLEYTKTQISLPKALQRYLLVREDLARAFAVQTSRYEFGLEFLP 300
QY 301 HLOTWVHNITRGKMGDAFHTNARDALNVSOAMWSLPEYKVSRLQKACGDAMNLGGR 360
DB 301 HLOTWVHNITRGKMGDAFHTNARDALNVSOAMWSLPEYKVSRLQKACGDAMNLGGR 360
QY 361 HVRSEQORNLIDLSTVTPEQIH 386
DB 361 HVRSEQORNLIDLSTVTPEQIH 386

```

```

RESULT 3
US-09-927-602-5
Sequence 5, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPLICANT: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 418
TYPE: PRT
ORGANISM: Mus musculus
US-09-927-602-5

```

```

Query Match 50.4%; Score 1028; DB 9; Length 418;
Best Local Similarity 56.6%; Pred. No. 8e-98; Indels 10; Gaps 6;
Matches 215; Conservative 43; Mismatches 112;

```

```

QY 1 MLPPK--KMKLLFLVSMALILAFPHMYSHNISLSKKAQPERMHVLYLSSWRSGSFFV 58
DB 25 MRLPSSSTVYMLSLAVQIGLVE--LVSRQVPS-SPAGLGERVHVLYLSSWRSGSFFV 80
QY 59 GOLFGHPDVYELMEPAHVMWTFKOSTAMMLHMAVRODLIRAVFLCDMSVDFDAYEPGPR 118
DB 81 GOLFGHPDVYELMEPAHVMWTFKOSTAMMLHMAVRODLIRAVFLCDMSVDFDAYEPGPR 139
QY 119 ROSSLFQWENSRLCASAPACDIIPODEIIPRAHCRLLCSQPFVEVEKACRSYSHVLEK 178
DB 140 NISDLFQWANSRLCSPYCEAFARONISSEVECKLCTRPGLAGQENCSYSHVLEK 199
QY 179 VREFNIQSLVPLKDPISLNIHVLVRDPAVFRSRETKGDLMDISRTVWGQHOEQLKK 238
DB 200 VREFNIQSLVPLKDPISLNIHVLVRDPAVFRSRETKGDLMDISRTVWGQHOEQLKK 258
QY 239 EDQPYVWQVQCSOLEYTKTQISLPKALQRYLLVREDLARAFAVQTSRYEFGLEFLP 298
DB 259 ADPLRLVNVQCSHVRIAEALHKPPFLQDRYRLVYEDLADPLVIRELYAFTGLGL 318
QY 299 LPHLOTWVHNITRGKMG--DHAFTNARDALNVSOAMWSLPEYKVSRLQKACGDAMNL 356
DB 319 TPOLOTWVHNITRGKMGDAFHTNARDALNVSOAMWSLPEYKVSRLQKACGDAMNL 378
QY 357 LGRVHVRSEQORNLIDLSTVTPEQIH 376
DB 379 LGRVHVRSEQORNLIDLSTVTPEQIH 398

```

```

RESULT 4
US-09-927-602-2
Sequence 2, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPLICANT: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 395

```


TYPE: PR1
ORGANISM: Homo Sapien
US-09-927-602-2

Query Match 50.0%; Score 1019; DB 9; Length 395;
Best Local Similarity 54.5%; Pred. No. 6.3e-97;
Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

QY 2 LPPKMKLLFLVYSOMAILALFFHMYSHNISLSMKAPERRMHVYVLSMRSSGSPVQOL 61
DB 14 LLLAOTCLLLFLVSRP-----GSSPAGEARVHVYVLSMRSSGSPVQOL 59
QY 62 FGOHPDVEYLMPEAMHVMTEFKOSTAMMLHMAVADLIRAVFLCDMSVDAVMEPGPRROS 121
DB 60 FNOHPDVEYLMPEAMHVMTEFKOSTAMMLHMAVADLIRAVFLCDMSVDAVMEPGPRROS 118
QY 122 SLFOWENSRALCSAPACDIITPQDEIIPRAHCRLLCSQOPEVEYKACRSYSHVYLKEVRF 181
DB 119 DLFOVAVSRALCSAPACDIITPQDEIIPRAHCRLLCSQOPEVEYKACRSYSHVYLKEVRF 178
QY 182 FNLQSLVPLKDPSSLNLIHVLVDRPRAVRSRERTGDLMDISRIYVGHQKLUKKEQD 241
DB 179 FNLQVLYPLSDPALNLIHVLVDRPRAVRSRERTGDLMDISRIYVGHQKLUKKEQD 237
QY 242 PYYVMOVYCOSOLEIYK--TIQSLPKALQERYLLVRYEDLARAPVQTSRMVEFYGLEFL 299
DB 238 GLRVYREVCSHVRIRAEATLKP--PPLRGYRLVREEDLARAPVQTSRMVEFYGLEFL 296
QY 300 PHLOTVHNITRGKMG--DHAFTNARDALNVQAMWSLPEYKVSRLQACGDAMNL 357
DB 297 POLKAMINITHGSGPGARREAFKTSRNALNVQAMWSLPEYKVSRLQACGDAMNL 356
QY 358 GYRHVRSQEOBNLLDILL-----STWVPEQIH 386
DB 357 GYRPVYSDEQRNALLDLVLPRLGNGFTWASTASH 392

RESULT 5

US-09-927-602-3
Sequence 3, Application US/09927602
Patent No. US20020061562A1

GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Methods of Treating Macular Corneal
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
PRIOR FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3
LENGTH: 395
TYPE: PR1
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: synthetic construct
NAME/KEY: VARIANT
LOCATION: (1)...(395)
OTHER INFORMATION: Xaa - any amino acid
US-09-927-602-3

Query Match 49.7%; Score 1013.5; DB 9; Length 395;
Best Local Similarity 55.9%; Pred. No. 2.4e-96;
Matches 212; Conservative 45; Mismatches 101; Indels 21; Gaps 6;

QY 2 LPPKMKLLFLVYSOMAILALFFHMYSHNISLSMKAPERRMHVYVLSMRSSGSPVQOL 61
DB 14 LLLAOTCLLLFLVSRP-----GSSPAGEARVHVYVLSMRSSGSPVQOL 59
QY 62 FGOHPDVEYLMPEAMHVMTEFKOSTAMMLHMAVADLIRAVFLCDMSVDAVMEPGPRROS 121

DB 60 FGOHPDVEYLMPEAMHVMTEFKOSTAMMLHMAVADLIRAVFLCDMSVDAVMEPGPRROS 118
QY 122 SLFOWENSRALCSAPACDIITPQDEIIPRAHCRLLCSQOPEVEYKACRSYSHVYLKEVRF 181
DB 119 DLFOVAVSRALCSAPACDIITPQDEIIPRAHCRLLCSQOPEVEYKACRSYSHVYLKEVRF 178
QY 182 FNLQSLVPLKDPSSLNLIHVLVDRPRAVRSRERTGDLMDISRIYVGHQKLUKKEQD 241
DB 179 FNLQVLYPLSDPALNLIHVLVDRPRAVRSRERTGDLMDISRIYVGHQKLUKKEQD 237
QY 242 PYYVMOVYCOSOLEIYK--TIQSLPKALQERYLLVRYEDLARAPVQTSRMVEFYGLEFL 299
DB 238 GLRVYREVCSHVRIRAEATLKP--PPLRGYRLVREEDLARAPVQTSRMVEFYGLEFL 296
QY 300 PHLOTVHNITRGKMG--DHAFTNARDALNVQAMWSLPEYKVSRLQACGDAMNL 357
DB 297 POLKAMINITHGSGPGARREAFKTSRNALNVQAMWSLPEYKVSRLQACGDAMNL 356
QY 358 GYRHVRSQEOBNLLDILL-----STWVPEQIH 386
DB 357 GYRPVYSDEQRNALLDLVLPRLGNGFTWASTASH 392

RESULT 6

US-09-927-602-4
Sequence 4, Application US/09927602
Patent No. US20020061562A1

GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Methods of Treating Macular Corneal
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
PRIOR FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4
LENGTH: 390
TYPE: PR1
ORGANISM: Homo Sapien
US-09-927-602-4

Query Match 49.5%; Score 1008; DB 9; Length 390;
Best Local Similarity 52.3%; Pred. No. 8.6e-96;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

QY 1 MLEPKMKLLFLVYSOMAILALFFHMYSHNISLSMKAPERRMHVYVLSMRSSGSPVQOL 60
DB 14 LLLAOTCLLLFLVSRP-----GSSPAGEARVHVYVLSMRSSGSPVQOL 59
QY 61 FGOHPDVEYLMPEAMHVMTEFKOSTAMMLHMAVADLIRAVFLCDMSVDAVMEPGPRROS 120
DB 60 FNOHPDVEYLMPEAMHVMTEFKOSTAMMLHMAVADLIRAVFLCDMSVDAVMEPGPRROS 118
QY 121 SLFOWENSRALCSAPACDIITPQDEIIPRAHCRLLCSQOPEVEYKACRSYSHVYLKEVRF 180
DB 119 DLFOVAVSRALCSAPACDIITPQDEIIPRAHCRLLCSQOPEVEYKACRSYSHVYLKEVRF 178
QY 181 FNLQSLVPLKDPSSLNLIHVLVDRPRAVRSRERTGDLMDISRIYVGHQKLUKKEQD 240
DB 179 FNLQVLYPLSDPALNLIHVLVDRPRAVRSRERTGDLMDISRIYVGHQKLUKKEQD 237
QY 241 QPYVMOVYCOSOLEIYK--TIQSLPKALQERYLLVRYEDLARAPVQTSRMVEFYGLEFL 298
DB 238 GLRVYREVCSHVRIRAEATLKP--PPLRGYRLVREEDLARAPVQTSRMVEFYGLEFL 296
QY 300 PHLOTVHNITRGKMG--DHAFTNARDALNVQAMWSLPEYKVSRLQACGDAMNL 356
DB 297 POLKAMINITHGSGPGARREAFKTSRNALNVQAMWSLPEYKVSRLQACGDAMNL 356

QY 357 LGYRHVRSQEQRLNLLDL-----STWYPE 383
 Db 357 LGYRPVYSADQQRDLTLVLPRGPDHFSWASPD 390

RESULT 7

US-09-927-602-8
 ; Sequence 8, Application US/09927602
 ; Patent No. US20020061562A1

GENERAL INFORMATION:

APPLICANT: Fukuda, Michiko N.
 APPLICANT: Akama, Tomoya O.
 TITLE OF INVENTION: Methods of Treating Macular Corneal
 TITLE OF INVENTION: Dystrophy
 FILE REFERENCE: P-LJ 4852
 CURRENT APPLICATION NUMBER: US/09/927,602
 CURRENT FILING DATE: 2001-08-09
 PRIOR APPLICATION NUMBER: US 09/638,211
 PRIOR FILING DATE: 2000-08-11

NUMBER OF SEQ ID NOS: 38
 SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 8

LENGTH: 171

TYPE: PRT

ORGANISM: Homo Sapien

US-09-927-602-8

Query Match 40.3%; Score 821; DB 9; Length 171;
 Best Local Similarity 60.0%; Pred. No. 6, 2e-77;
 Matches 171; Conservative 0; Mismatches 0; Indels 114; Gaps 2;

QY 32 SSSLSMAQPEPRMHLVLSRRSSSFVGLFGQHPVFLMEPRMHWMTFKOSTAMMLH 91
 Db 1 SSSLSMAQPEPRMHLVLSRRSSSFVGLFGQHPVFLMEPRMHWMTFKOSTAMMLH 52
 QY 92 MAVRDLIRAVFLCDMSVPAVMEPRRQSSLFQWENSRLCSAPACDIIPDEIIPRAH 151
 Db 53 ----- 52
 QY 152 CRLLCSQOPPEVVEKACRSYSHVLEKVRFNLSLPLKDPNLHLVHLVDRPRAVE 211
 Db 53 -----KACRSYSHVLEKVRFNLSLPLKDPNLHLVHLVDRPRAVE 98
 QY 212 RSRRTKGLMIDSRVMOGHOKLKEKEDPYTVMOYICOSLEIKTIOSLPKALOERY 271
 Db 99 RSRRTKGLMIDSRVMOGHOKLKEKEDPYTVMOYICOSLEIKTIOSLPKALOERY 126
 QY 272 LTVRYEDLARAPVQTSRMTEFVGLFPLHLOTWVHNITRGKMG 316
 Db 127 LTVRYEDLARAPVQTSRMTEFVGLFPLHLOTWVHNITRGKMG 171

RESULT 8

US-10-212-933-2
 ; Sequence 2, Application US/10212933
 ; Publication No. US20030008366A1

GENERAL INFORMATION:

APPLICANT: Uchimura, Kenji
 APPLICANT: Muramatsu, Hideki
 APPLICANT: Kadomatsu, Kenji
 APPLICANT: Kannagi, Reiji
 APPLICANT: Habuchi, Osami
 APPLICANT: Muramatsu, Takashi
 TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-
 TITLE OF INVENTION: SULFOTRANSFERASE AND
 TITLE OF INVENTION: DNA ENCODING THE SAME
 FILE REFERENCE: TOYAMA1.001AUS
 CURRENT APPLICATION NUMBER: US/10/212,933
 CURRENT FILING DATE: 2002-08-05
 PRIOR APPLICATION NUMBER: US/09/263,023
 PRIOR FILING DATE: 1999-03-05
 PRIOR APPLICATION NUMBER: JP 10-54007

;; PRIOR FILING DATE: 1998-03-05
 ;; PRIOR APPLICATION NUMBER: JP 10-177844
 ;; PRIOR FILING DATE: 1998-06-24
 ;; NUMBER OF SEQ ID NOS: 10
 ;; SOFTWARE: FastSeq for Windows Version 3.0
 ;; SEQ ID NO 2
 ;; LENGTH: 483
 ;; TYPE: PRT
 ;; ORGANISM: Mus musculus
 ;; US-10-212-933-2

Query Match 29.4%; Score 598.5; DB 15; Length 483;
 Best Local Similarity 36.8%; Pred. No. 2, 9e-53;
 Matches 134; Conservative 64; Mismatches 137; Indels 29; Gaps 6;

QY 41 ERMHVLSWRSSSFVGLFGQHPVFLMEPRMHWMTFKOSTAMMLHVAVRDLIRA 100
 Db 116 KRQLYVFTWRSSSFVGLFGQHPVFLMEPRMHWMTFKOSTAMMLHVAVRDLIRA 175
 QY 101 VFLCDMSVPAVMEPRRQSSSFVGLFGQHPVFLMEPRMHWMTFKOSTAMMLH 155
 Db 176 LYRCDLSVFLQYSPASGGRNLTTLGIFGAATKVKVCSPLCPAY-RKEVGLVDDRVCK 234
 QY 156 -CSQOPPEVVEKACRSYSHVLEKVRFNLSLPLKDPNLHLVHLVDRPRAVE 214
 Db 235 KCPQRLARPEECRKYRTVILKGVRFVAVLAPLKDPLDKVHLVDRPRAVSSR 294
 QY 215 ERTKGLMIDSRVMOGHOKLKEKEDPYTVMOYICOSLEIKTIOSLPKALOERY 252
 Db 295 IRRHLLRESLQVNSRDPRAHMPLEAGHKLAKKKGCGPADYHALGAMETICS 354
 QY 253 QLEIYTIOSLPALDERILVYEDLARAPVQTSRMTEFVGLFPLHLOTWVHNITRG 312
 Db 355 MARTLOTALQPPMLQGHVLYVEDLVDPVTTLRRVYDFGLVSPMEQFALMTSG 414
 QY 313 KGMGDHAFHNMARDALNVSAWMSLPYEKVSRLKACGDAMMLGRRHVSQEQRLN 372
 Db 415 SGSSKPEFVYSARNATQANAMNTALTPOIKQVEEFCYOPNAVLOERYNSPEEVKIDS 474
 QY 373 LDLL 376
 Db 475 KTL 478

RESULT 9

US-10-212-933-4
 ; Sequence 4, Application US/10212933
 ; Publication No. US20030008366A1

GENERAL INFORMATION:

APPLICANT: Uchimura, Kenji
 APPLICANT: Muramatsu, Hideki
 APPLICANT: Kadomatsu, Kenji
 APPLICANT: Kannagi, Reiji
 APPLICANT: Habuchi, Osami
 APPLICANT: Muramatsu, Takashi
 TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-
 TITLE OF INVENTION: SULFOTRANSFERASE AND
 TITLE OF INVENTION: DNA ENCODING THE SAME
 FILE REFERENCE: TOYAMA1.001AUS
 CURRENT APPLICATION NUMBER: US/10/212,933
 CURRENT FILING DATE: 2002-08-05
 PRIOR APPLICATION NUMBER: US/09/263,023
 PRIOR FILING DATE: 1999-03-05
 PRIOR APPLICATION NUMBER: JP 10-54007
 PRIOR FILING DATE: 1998-03-05
 PRIOR APPLICATION NUMBER: JP 10-177844
 PRIOR FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 10
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 4
 LENGTH: 484
 TYPE: PRT
 ORGANISM: Homo sapiens

US-10-212-933-4

Query Match	28.8%;	Score 587.5;	DB 15;	Length 484;
Best Local Similarity	36.0%;	Pred. No. 4e-52;		
Matches 131;	Conservative 67;	Mismatches 137;	Indels 29;	Gaps 6;

```

QY      41  ERNHNVLVLSNMRSGSFFVGLFGOHDPVEYLMEPRAHNMHMTFESTAMLMIAVRDLIRA 1000
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dh      117  KRHMIVLVTTHMSGSSFFBSELNQNPEVFFLEPRHNVHWOXLYLPEDAVSLQAGARDMLISA 1768
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      101  VFLCDMSVDEAYMEBPGRQS---SLFQWENSRAIACSAPACDIIPODEILPRAHCRL- 1558
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dh      177  LYRCDSLVFQIVSPAGSGRNLTLTGIGAAITNKVYCSPLCPAY-KREYVGLVDDBYCK 2359
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      156  -CSQGFVEYERKACRSYSHVYLKEYRFRNLOSLYLKDPSLNIHYHVLNDRPRAVRSR 2144
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dh      236  KCPQQLARFEECECRKRYLTIKGRVFDVAVLATLPDRLDLKVLHVLADPRAVASSR 2959
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      215  ERFGSLMIDSRIVM-----GQHEQKLKE-----DQRY---VWQVYCS 25
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dh      296  IRRHQLINRESLQVYNSRDPRAHRMPELEAGHKLGAKEEGVGADYHIALGAEVICS 3555
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      253  QLEIYVLTOSLPKALOERYLVREYEDLARAPAYQTSRNKVEYVGEFLPHLOTWVHNITRG 3142
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dh      356  MANTLQTLAQPPDWLQGHLYLVREYDVGDPYKTLIRRYDVGGLVSPENMQFRLMNTSG 415
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      313  KGMGDHAFHTNADALNLSQANRMSLPYEKVSRLQACGDAMNLGYRHVNSQEQQBNLL 3737
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dh      416  SGSSSKRPFVVSARNATQANAMRATLTLEQILQVVEEFCYCPMAVLYGERVNSPEVXKDS 4757
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      373  LDLL 376
      : :
Dh      476  KTL 479

```

```

RESULT 10
US-09-833-790-255
; Sequence 255, Application US/09833790
; Patient No. US20020068288A1
GENERAL INFORMATION:
; APPLICANT: Lodges, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Secretist, Heather
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Indirlas, Carol Y.
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LONG CANCER
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 255
; LENGTH: 531
; TYPE: PRF
; ORGANISM: Homo sapiens
; US-09-833-790-255

```

[illegible]

```

Db      283 KCPQRRLAREECCRCRYRLVYIKGVAFVDAVLPLRLNRPALDLVHILVNDPRVAVSR 344
OY      215 ERTKGDLMDLSRYM-----GHEQKKEKKE----DQPY---VMQYICQ 257
Db      343 IRSHHGIRISLQVVSRRDPRAHRMDFLEAAGHKLGAKKEGVGPADYHALGAEVION 402
OY      253 QLEHYKTIQSLPRLQDEYLLVREYEDLAPVPAQTSRMVEFVGLPEFLPHLOTVMYINIRG 314
Db      403 MARTLQRLALQPRPMDLOGHLYVREYEDVDPRKTLRYVDYDFGLLVSPEDMFALMTMSG 466
OY      313 KGMGDRAFTHTNDALNVSOAMRWSLPYEKVSRBLQACGDANMLLGYRHVNSDEQRMNL 374
Db      463 SGSSSKPEFVVASNNATQANAMWTALTFOOIKOVEEFCYQPAVYLVGERVNSPEBEVKDL 522
OY      373 LDLL 376
Db      523 KTL 526

```

```

RESULT 11
US-09-927-602-6
: Sequence 6, Application US/09927602
: Patent No. US20020061562A1
: GENERAL INFORMATION:
: APPLICANT: Fukuda, Michiko N.
: APPLICANT: Akama, Tomoya O.
: TITLE OF INVENTION: Methods of Treating Macular Corneal
: TITLE OF INVENTION: Dystrophy
: FILE REFERENCE: P-IJ 4852
: CURRENT APPLICATION NUMBER: US/09/927,602
: CURRENT FILING DATE: 2001-08-09
: PRIOR APPLICATION NUMBER: US 09/638,211
: PRIOR FILING DATE: 2000-08-11
: NUMBER OF SEQ ID NOS: 38
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 6
: LENGTH: 169
:
: TYPE: PR1
: ORGANISM: Homo Sapien
US-09-927-602-6

```

[illegible]

RESULT 12
US-09-927-602-7
: Sequence 7, Application US/09927602
: Patent No. US20020061562A1
: GENERAL INFORMATION:
: APPLICANT: Fukuda, Michiko N.
: APPLICANT: Akama, Tomoyo O.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 11:06:37 ; Search time 6872.26 Seconds

(Without alignments)
11465.210 Million cell updates/sec

Title: US-09-645-078-3

Perfect score: 1926
Sequence: 1 gggcactactactacatctg.....ccctgcacatcttcttaag 1926

Scoring table:

IDENTITY_NIC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
GenBank:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_in:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htg_hum:*
40: em_htg_mus:*
41: em_htg_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1926	100.0	1926	6 BD134773	BD134773 GLYCOSYL
2	1926	100.0	1926	10 AF131236	AF131236 Mus muscu
3	1481.2	76.9	2201	10 AF109135	AF109135 Mus muscu
C 4	1277	66.3	232745	2 AC130174	AC130174 Rattus no
C 5	1047	54.4	262679	2 AC110314	AC110314 Rattus no
C 6	753.2	39.1	183228	2 AC10547	AC10547 Homo sapi
C 7	753.2	39.1	206943	2 AC138848	AC138848 Homo sapi
C 8	751.2	39.0	1333	9 AF149783	AF149783 Homo sapi
9	743.2	38.6	1992	9 AF280088	AF280088 Homo sapi
10	743.2	38.6	2011	9 AK026635	AK026635 Homo sapi
11	743.2	38.6	2037	9 BC035282	BC035282 Homo sapi
12	741.6	38.5	1979	6 BD127258	BD127258 Homo sapi
13	741.6	38.5	1979	9 AK074746	AK074746 Homo sapi
14	741.6	38.5	2032	6 AR203335	AR203335 Sequence
15	741.6	38.5	2032	6 BD134772	BD134772 GLYCOSYL
16	741.6	38.5	2032	9 AF131235	AF131235 Homo sapi
17	717	37.2	2065	6 AR310485	AR310485 Sequence
18	468.8	24.3	877	6 BD124832	BD124832 Primer fo
19	468.8	24.3	877	6 BD126950	BD126950 primer fo
20	370.6	19.2	1647	6 AX327330	AX327330 Sequence
21	370.6	19.2	2544	9 AF219990	AF219990 Homo sapi
22	370.6	19.2	3786	9 AF280086	AF280086 Homo sapi
23	370.6	19.2	71503	9 AF219991	AF219991 Homo sapi
C 24	370.6	19.2	157358	9 AC009163	AC009163 Homo sapi
C 25	370.6	19.2	208185	2 AC009105	AC009105 Homo sapi
26	365.2	19.0	229007	2 AC095664	AC095664 Rattus no
27	365.2	19.0	245248	2 AC117869	AC117869 Rattus no
28	355.6	18.5	1740	10 AF176841	AF176841 Mus muscu
29	355.6	18.5	1989	10 AF176840	AF176840 Mus muscu
30	355.6	18.5	159170	2 AC115914	AC115914 Mus muscu
31	355.6	18.5	209198	2 AC127315	AC127315 Mus muscu
32	354	18.4	1462	9 AF176839	AF176839 Homo sapi
33	354	18.4	2170	9 AF176838	AF176838 Homo sapi
34	354	18.4	3278	9 AF246718	AF246718 Homo sapi
35	354	18.4	194832	9 AC025287	AC025287 Homo sapi
C 36	122.6	6.4	209724	2 BX322552	BX322552 Danilo rer
C 37	116	6.0	1654	10 AB008937	AB008937 Mus muscu
C 38	116	6.0	7515	10 AB062107	AB062107 Mus muscu
C 39	116	6.0	223267	2 AC121503	AC121503 Mus muscu
40	114	5.9	1241	9 AF090137	AF090137 Homo sapi
41	114	5.9	1458	6 E16306	E16306 cDNA encodl
42	114	5.9	2190	6 AR211733	AR211733 Sequence
43	114	5.9	2190	6 HS065637	HS065637 Homo sapien
44	114	5.9	2415	9 AB003791	AB003791 Homo sapi
45	114	5.9	2453	9 BC028235	BC028235 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS BD134773 1926 bp DNA linear PAT 18-SEP-2002
DEFINITION GLYCOSYLsulfoltransferase-3.
ACCESSION BD134773
VERSION BD134773.1 GI:23229718
KEYWORDS UP 2002507409-A/2.
SOURCE Mus sp.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1926)
AUTHORS Bistup,A., Rosen,S.D., Tangemann,K. and Hemmerlich,S.
TITLE Glycosylsulfoltransferase-3
JOURNAL Patent: JP 2002507409-A 2 12-MAR-2002;

COMMENT THE REGENTS OF THE UNIVERSITY OF CALIFORNIA, SYNTAX INC

OS Mus sp. (mouse)
 PN JP 2002507409-A/2
 PD 12-MAR-2002
 PF 26-FEB-1999 JP 2000537979
 PR 20-MAR-1998 US 09/045284,12-NOV-1998 US 09/190911 PI
 ANNETTE BISTRUP, STEVEN D ROSEN, KIRSTEN TANGEMANN, STEFAN PI
 HEMMERICH
 PC C12N9/10, A01K67/027, A61K45/00, A61K48/00, A61P29/00, A61P37/06,
 PC A61P43/00,
 PC C12N1/15, C12N1/21, C12N5/10, C12N15/09, C12Q1/48, C12N5/00, C12N15/09
 CC Glycosylsulfotransferase-3
 FH key Location/Qualifiers
 FT source 1.1926
 FT Location/Qualifiers
 1.1926 /organism="Mus sp. (mouse)"
 /mol_type="genomic DNA"
 /db_xref="taxon:10095"

BASE COUNT 426 a 520 c 484 g 496 t

ORIGIN

Query Match 100.0%; Score 1926; DB 6; Length 1926;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1926; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GGGCATCTAACCTTACACTGTGTGAGACAGACAGCTTGTGCTACAAAGGCGACACCTGT 60
 1 GGGCATCTAACCTTACACTGTGTGAGACAGACAGCTTGTGCTACAAAGGCGACACCTGT 60
 61 GTGAGGGGTGTAGAAAGGTGTGGGTGTGGCAGAACTCCCTATAGTAAATGTGCTG 120
 61 GTGAGGGGTGTAGAAAGGTGTGGGTGTGGCAGAACTCCCTATAGTAAATGTGCTG 120
 121 GGTAGGATATCTCGGTGTGTGATGAGTAAAGAACCCAGAGGGGATTTTAAAGACT 180
 121 GGTAGGATATCTCGGTGTGTGATGAGTAAAGAACCCAGAGGGGATTTTAAAGACT 180
 121 GGTAGGATATCTCGGTGTGTGATGAGTAAAGAACCCAGAGGGGATTTTAAAGACT 180
 181 TGTACATAGATGAGTATGATGATTAAGAGCCAGAACTTGTGACAGAGGATCTGGA 240
 181 TGTACATAGATGAGTATGATGATTAAGAGCCAGAACTTGTGACAGAGGATCTGGA 240
 181 TGTACATAGATGAGTATGATGATTAAGAGCCAGAACTTGTGACAGAGGATCTGGA 240
 241 CAATACCTTGCAGAGATGTTTGGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 300
 241 CAATACCTTGCAGAGATGTTTGGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 300
 301 CTGGATGGGATTCAGAGAGAGCCGAGAGGTAGTGTGATACAACTTAACCTACAGCCCA 360
 301 CTGGATGGGATTCAGAGAGAGCCGAGAGGTAGTGTGATACAACTTAACCTACAGCCCA 360
 361 TCCCTCTGCTTCTCTTCAAGGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
 361 TCCCTCTGCTTCTCTTCAAGGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
 421 AAGGAGGCTGTGATGTTCTGTGGTTCAGAGTATGATGATGATGATGATGATGATGAT 480
 421 AAGGAGGCTGTGATGTTCTGTGGTTCAGAGTATGATGATGATGATGATGATGATGAT 480
 481 TGTGCTTCCAGACAGACCTTTCAGAGAGGAGAGTCCAGAGAGCCCGTGCATGTGCTG 540
 481 TGTGCTTCCAGACAGACCTTTCAGAGAGGAGAGTCCAGAGAGCCCGTGCATGTGCTG 540
 541 TGTGCTTCTCTGAGGCTGAGAGCTCTTCTTGTGGAGAGCTTTCGGGAGAGACCCGG 600
 541 TGTGCTTCTCTGAGGCTGAGAGCTCTTCTTGTGGAGAGCTTTCGGGAGAGACCCGG 600
 601 ATGTGTTTACCTGATGAGAGCTGTGCTGTGATGTGATGATGATGATGATGATGATGAT 660
 601 ATGTGTTTACCTGATGAGAGCTGTGCTGTGATGTGATGATGATGATGATGATGATGAT 660
 661 CCTGAGAGCTGACATGGCTGTGGGAGATCTTCTCGTTCCTCTCTGATGACATGA 720

661 CCTGAGAGCTGACATGGCTGTGGGAGATCTTCTCGTTCCTCTGATGACATGA 720
 721 GGGTCTTATATCCCTATGAGAACCCAGAGCCCGGAGAGAGTCCAGCTTCCAGTGG 780
 721 GGGTCTTATATCCCTATGAGAACCCAGAGCCCGGAGAGAGTCCAGCTTCCAGTGG 780
 781 AGCAAGAGCCGAGCCCTGTGCTGAGAGCCCTGTGCTGAGTCTTCTCTGAGAGATGA 840
 781 AGCAAGAGCCGAGCCCTGTGCTGAGAGCCCTGTGCTGAGTCTTCTCTGAGAGATGA 840
 841 GGTACCCCAAGCAGTCAAGCTGTGCTGAGAGCCCTGTGCTGAGTCTTCTCTGAGAG 900
 841 GGTACCCCAAGCAGTCAAGCTGTGCTGAGAGCCCTGTGCTGAGTCTTCTCTGAGAG 900
 901 CCTGAGAGCTGACATGGCTGTGGGAGATCTTCTCGTTCCTCTGATGACATGA 960
 901 CCTGAGAGCTGACATGGCTGTGGGAGATCTTCTCGTTCCTCTGATGACATGA 960
 961 TGTATCCTACTTCAAGAGCCCTTCCCTCAAGCTGAGTCAAGCTGAGTCAAGAG 1020
 961 TGTATCCTACTTCAAGAGCCCTTCCCTCAAGCTGAGTCAAGCTGAGTCAAGAG 1020
 1021 CCCGGGCGGTGTTCCGATCCCGGAGACACCACTTAAGTCAAGTCAAGTCAAG 1080
 1021 CCCGGGCGGTGTTCCGATCCCGGAGACACCACTTAAGTCAAGTCAAGTCAAG 1080
 1081 TGTGCTAGGAGAGATTTGGAAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
 1081 TGTGCTAGGAGAGATTTGGAAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
 1141 AGATCATCTGCAAAAGCCAGGTGACATAGTCAAGAGCCCTCTGAGAGCTC 1200
 1141 AGATCATCTGCAAAAGCCAGGTGACATAGTCAAGAGCCCTCTGAGAGCTC 1200
 1201 TGTGCTAGGAGAGATTTGGAAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 1201 TGTGCTAGGAGAGATTTGGAAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 1261 CGACAGAGATTAATTAATTTGTGGGTTGATTTTGTGGGTTGATTTTGTGGGTTG 1320
 1261 CGACAGAGATTAATTAATTTGTGGGTTGATTTTGTGGGTTGATTTTGTGGGTTG 1320
 1321 ACAATGTCAACCCGAG 1380
 1321 ACAATGTCAACCCGAG 1380
 1381 CCTCAAGCTCTCAAG 1440
 1381 CCTCAAGCTCTCAAG 1440
 1441 AAGATGCTGCTGAG 1500
 1441 AAGATGCTGCTGAG 1500
 1501 AACAAAGCAAGCTGTGCTGAGATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1560
 1501 AACAAAGCAAGCTGTGCTGAGATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1560
 1561 GAGAAGGTTAAG 1620
 1561 GAGAAGGTTAAG 1620
 1621 ACAGAAAGCTTAAGATTAACCAAGTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1680
 1621 ACAGAAAGCTTAAGATTAACCAAGTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1680
 1681 GGTCTTGTGTATATCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
 1681 GGTCTTGTGTATATCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
 1741 ATCTTGTGTGTGAAATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
 1741 ATCTTGTGTGTGAAATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800

Db 1741 ATCTTCTGCTGAAAACTAGAACCTTAAGACAGCCCTTGAACCTGTCAAGCAAG 1800
 Qy 1801 ACCTTCTGTAACCTTGCCCTTCTTACCTGATACCTTGAGACAGCGGCTGAGCA 1860
 Db 1801 ACTTCTTGTACCTTGCCCTTCTTACCTGATACCTTGAGACAGCGGCTGAGCA 1860
 Qy 1861 TACTGACACAGCAACAGCATCTGTGAGTGTCTGTAAACCTCCCTGTCAATCTTT 1920
 Db 1861 TACTGACACAGCAACAGCATCTGTGAGTGTCTGTAAACCTCCCTGTCAATCTTT 1920
 Qy 1921 TCTAAG 1926
 Db 1921 TCTAAG 1926
 RESULT 2
 AF131236 1926 bp DNA linear ROD 10-AUG-1999
 LOCUS
 DEFINITION Mus musculus N-acetylglucosamine 6-O-sulfotransferase gene,
 complete cds.
 ACCESSION AF131236
 VERSION AF131236.1 GI:4927115
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM
 REFERENCE
 AUTHORS Huang, C.C., Kannagi, R., Rosen, S.D. and Hemmerich, S.
 Bistup, A., Bhakta, S., Lee, J.K., Belov, Y.Y., Gunn, M.D., Zuo, F.R.,
 Huang, C.C., Kannagi, R., Rosen, S.D. and Hemmerich, S.
 Sulfotransferases of two specificities function in the
 reconstitution of high endothelial cell ligands for L-selectin
 J. Cell Biol. 145 (4), 899-910 (1999)
 JOURNAL 99264336
 MEDLINE 10330415
 PUBMED
 REFERENCE 2 (bases 1 to 1926)
 AUTHORS Bistup, A., Tangemann, K., Bhakta, S., Lee, J.-K., Belov, Y.Y.,
 Gunn, M.D., Zuo, F.-R., Huang, C.-C., Kannagi, R., Rosen, S.D. and
 Hemmerich, S.
 Direct Submission
 Submitted (25-FEB-1999) Respiratory Diseases, Roche Bioscience,
 3401 Hillview Avenue, Palo Alto, CA 94304, USA
 TITLE
 JOURNAL Location/Qualifiers
 FEATURES
 source
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /cell_type="high endothelial"
 /tissue_type="tonsil"
 <405..>1571
 /product="N-acetylglucosamine 6-O-sulfotransferase"
 405..1571
 /function="biosynthesis of endothelial ligands for
 L-selectin"
 /note="HEC-G1CMAC-6-ST"
 /codon_start=1
 /product="N-acetylglucosamine 6-O-sulfotransferase"
 /protein_id="FAD3016.1"
 /db_xref="GI:4927116"
 /translation="MMLKRGRLMFLGSOVIVVAFIHSVHRHLSQRESRRPVH
 LVLSWRSGSSFYGLFGHDPVFEYLMPEAMHWMPTSSAMKLHAAVDLRSVL
 CDMSEVPAVYNNPGKROSLFQWOSRALSAPYCDFFPHEIISPHCKLGGQDF
 DWYEKASHRGFYVLKEVRLSLOALYPLTDPSLNHVHVIVDPPVAPRSREHTTI
 ELWDVSHVIGHGLFTKKEEDOPYAKIICKSVDYVKAICITLPEALDORYFLRIE
 DLYRPLAQTRILKRYGLDPLPHLQTVNVTIRGKMGQHAFTNARNLNVSAWR
 WSLPEYVSDLDGCGEAMDLGLIYVRSQEQGNLSLDSLSSHLIGOVFREG"
 BASE COUNT 426 a 520 c 484 g 496 t
 ORIGIN
 Query Match 100.0%; Score 1926; DB 10; Length 1926;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1926; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCATCTAATCTACCTTGCTGAGACAGCAAGCTTTGGCTACAAAGGCCACACTCT 60
 Db 1 GGCATCTAATCTACCTTGCTGAGACAGCAAGCTTTGGCTACAAAGGCCACACTCT 60
 Qy 61 GTCAGGGGTGTAGAAAGGTGTGGGTGTGCGAGAATCCCTATAGTATTAATGTGTG 120
 Db 61 GTCAGGGGTGTAGAAAGGTGTGGGTGTGCGAGAATCCCTATAGTATTAATGTGTG 120
 Qy 121 GGTAGATATTTCTGGTGTGTTGATGATGAGAAAGCCAGAGGATGATTAAAGACT 180
 Db 121 GGTAGATATTTCTGGTGTGTTGATGATGAGAAAGCCAGAGGATGATTAAAGACT 180
 Qy 181 TGTAACTAGAAATCAGATGATCAATTAAGAGCCAGAAATTTCTTGAGAGGATTTGGA 240
 Db 181 TGTAACTAGAAATCAGATGATCAATTAAGAGCCAGAAATTTCTTGAGAGGATTTGGA 240
 Qy 241 CAATTAATCTGAGGAATGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 300
 Db 241 CAATTAATCTGAGGAATGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 300
 Qy 301 CTTGATGGAATCCAGAAAGCCGAAAGTATGCTGTAACTAATCTACAGCCCA 360
 Db 301 CTTGATGGAATCCAGAAAGCCGAAAGTATGCTGTAACTAATCTACAGCCCA 360
 Qy 361 TCCCTGCTTGGCTTCTTCAAGGCTTCTTCCGAGAGATGATGCTGTGAAGA 420
 Db 361 TCCCTGCTTGGCTTCTTCAAGGCTTCTTCCGAGAGATGATGCTGTGAAGA 420
 Qy 421 AAGGAGGCTGCTATGTTCTCGGTTCCAGGTATGTTGATGCTTCTTCAATCA 480
 Db 421 AAGGAGGCTGCTATGTTCTCGGTTCCAGGTATGTTGATGCTTCTTCAATCA 480
 Qy 481 TGTCCGTCCACAGACACCTTTCCAGAGGAGAGTCCAGAGGCCGCTGATGCTGG 540
 Db 481 TGTCCGTCCACAGACACCTTTCCAGAGGAGAGTCCAGAGGCCGCTGATGCTGG 540
 Qy 541 TGTCTGTTCCCTGCGGTCAGAGATCTCTTTTGGGACAGCTTTTGGGACAGCCGG 600
 Db 541 TGTCTGTTCCCTGCGGTCAGAGATCTCTTTTGGGACAGCTTTTGGGACAGCCGG 600
 Qy 601 ATGTGTTCTACCTGATGAGGCTGCTGCGATGTTGATGATCAACAGCAGCAGAG 660
 Db 601 ATGTGTTCTACCTGATGAGGCTGCTGCGATGTTGATGATCAACAGCAGCAGAG 660
 Qy 661 CCTGAGGCTGCAATGCTGTGGGATCTTTCGCTTCCGTCTGCTGATACATGA 720
 Db 661 CCTGAGGCTGCAATGCTGTGGGATCTTTCGCTTCCGTCTGCTGATACATGA 720
 Qy 721 GCGTCTTGTGATGCTTACATGAACCCAGGCCCCGGAACAAGTCCAGGCTTCCAGTGGG 780
 Db 721 GCGTCTTGTGATGCTTACATGAACCCAGGCCCCGGAACAAGTCCAGGCTTCCAGTGGG 780
 Qy 781 AGCAAGAGCGGCGCTGTGCTCAGCGCTGTGTGTGACTTCTCCCTGCGCAGAGATCA 840
 Db 781 AGCAAGAGCGGCGCTGTGCTCAGCGCTGTGTGTGACTTCTCCCTGCGCAGAGATCA 840
 Qy 841 GCTACCCCAACGACTGCAAGTCTCTGCGGCTGAGCAGCCCTTGTGATGATGATGAAGG 900
 Db 841 GCTACCCCAACGACTGCAAGTCTCTGCGGCTGAGCAGCCCTTGTGATGATGATGAAGG 900
 Qy 901 CCTGCGCTCTCAGAGGCTTCTGTACTCAAGAGAGTGCCTTCTCAGCTGCAAGGCC 960
 Db 901 CCTGCGCTCTCAGAGGCTTCTGTACTCAAGAGAGTGCCTTCTCAGCTGCAAGGCC 960
 Qy 961 TCTATCACTACTACAGGACCTTCTCCTCAACTGACAGCTGTGACCTGTCGAGAGC 1020
 Db 961 TCTATCACTACTACAGGACCTTCTCCTCAACTGACAGCTGTGACCTGTCGAGAGC 1020
 Qy 1021 CCGGGGCGGTGCTCGATCCCGGAGACACACCATATGAATCATGTTACAGTATA 1080
 Db 1021 CCGGGGCGGTGCTCGATCCCGGAGACACACCATATGAATCATGTTACAGTATA 1080

```

OY 1081 TTGTGCTAGGAGCATTTTGAAGATGATGAGAGAGAACAGCCCTTTATGCCATGA 1140
Db 1081 TTGTGCTAGGAGCATTTTGAAGATGATGAGAGAGAACAGCCCTTTATGCCATGA 1140
OY 1141 AGATCATCTGCAAAAGCCAGTGTGACATAGTCAAGGCCATCCAAACCCCTCGAAGCTC 1200
Db 1141 AGATCATCTGCAAAAGCCAGTGTGACATAGTCAAGGCCATCCAAACCCCTCGAAGCTC 1200
OY 1201 TGCAGAGCGGTACCTGTTCTGAGATGATGAGACCTGGTTCGGGACCCCTGGCCACA 1260
Db 1201 TGCAGAGCGGTACCTGTTCTGAGATGATGAGACCTGGTTCGGGACCCCTGGCCACA 1260
OY 1261 GCACAGACTATATTAATTTGGGGTGGATTTTTTGGCCCAACCTCCCAAAACATGGGTTT 1320
Db 1261 GCACAGACTATATTAATTTGGGGTGGATTTTTTGGCCCAACCTCCCAAAACATGGGTTT 1320
OY 1321 ACAATGTCAACCCGCGCAAGGCGATGAGTGCAGCATGCTTCCATACTATAAGCAGAGAG 1380
Db 1321 ACAATGTCAACCCGCGCAAGGCGATGAGTGCAGCATGCTTCCATACTATAAGCAGAGAG 1380
OY 1381 CCCGCAAGCTCTCTGAGCGGTGGCTTCCCTTACGTAAGAAAAGTTTCCAGCTTC 1440
Db 1381 CCCGCAAGCTCTCTGAGCGGTGGCTTCCCTTACGTAAGAAAAGTTTCCAGCTTC 1440
OY 1441 AAGATGCGGTGCGGTAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
Db 1441 AAGATGCGGTGCGGTAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
OY 1501 AACAGAGCAACCTGTCCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
Db 1501 AACAGAGCAACCTGTCCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
OY 1561 GAGAAAGTTAAGAGAGTGTCTGCAACCCCTGGTTCAGGCTTACGTAAGCAATTAACGC 1620
Db 1561 GAGAAAGTTAAGAGAGTGTCTGCAACCCCTGGTTCAGGCTTACGTAAGCAATTAACGC 1620
OY 1621 ACAGAAAGCTTAAAGATATACCAAACTGATGATGATGATGATGATGATGATGATGATGAT 1680
Db 1621 ACAGAAAGCTTAAAGATATACCAAACTGATGATGATGATGATGATGATGATGATGATGAT 1680
OY 1681 GGTCTTTGTGTATATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 1740
Db 1681 GGTCTTTGTGTATATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 1740
OY 1741 ATCTTTCTGTCTTGAATAATCTATTAAGAACCTTAAGACGCCCTTGGACCTGTCAAGCA 1800
Db 1741 ATCTTTCTGTCTTGAATAATCTATTAAGAACCTTAAGACGCCCTTGGACCTGTCAAGCA 1800
OY 1801 ACTTCTGTATACCTTGGCTCTTACCTGATGATGATGATGATGATGATGATGATGATGATGAT 1860
Db 1801 ACTTCTGTATACCTTGGCTCTTACCTGATGATGATGATGATGATGATGATGATGATGATGAT 1860
OY 1861 TACTGAGACAGCAAAAGCATCTGTGAGATGTGTCTGTAACCTCCCTGTGACATCTTT 1920
Db 1861 TACTGAGACAGCAAAAGCATCTGTGAGATGTGTCTGTAACCTCCCTGTGACATCTTT 1920
OY 1921 TCTAAG 1926
Db 1921 TCTAAG 1926

```

```

RESULT 3
AF109155 2201 bp mRNA linear ROD 27-JUL-1999
LOCUS Mus musculus l-selectin ligand sulfotransferase mRNA, complete cds.
DEFINITION
AF109155
ACCESSION
AF109155.1 GI:5596405
VERSION
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 2201)

```

```

AUTHORS
Hirooka, N., Petryniak, B., Nakayama, J., Tsuboi, S., Suzuki, M.,
Yeh, J.C., Izawa, D., Tanaka, T., Miyasaka, M., Lowe, J.B. and Fukuda, M.
TITLE
A novel, high endothelial venule-specific sulfotransferase
expresses 6-sulfo sialyl Lewis(x), an L-selectin ligand displayed
by CD34
JOURNAL
Immununity 11 (1), 79-89 (1999)
MEDLINE
99361934
PUBMED
10435581
REFERENCE
2 (bases 1 to 2201)
AUTHORS
Hirooka, N. and Fukuda, M.
TITLE
Direct Submission
JOURNAL
Submitted (23-NOV-1998) Glycobiology, The Burnham Institute, 10901
North Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES
source
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
319..1485
/function="directs expression of 6-sulfo sialyl Lewis x,
an L-selectin ligand displayed by CD34"
/note="LSSR"
/codon_start=1
/product="L-selectin ligand sulfotransferase"
/protein_id="A045579.1"
/db_xref="GI:5596406"
/translation="MMLKKGRILNLFSGQIVVALFEHMSVHRHLSORESRPVH
LVLSWRSSSFVQDLFGQHPDVFYLMEMAMVMTFTSTAMKLMAVRDLRSVFL
CMSEVDAMNNGPKRSLPQWEDSRLSCSPVDFEHSISPRKHLKLGQOFP
DMVEKARSHGFPVILKEVRFSLQALYPLPSLNIHVYHVRPFRREHRTT
ELVYSHIYLGQHLFETKEEDPYTAMKLTICSYQDIYKAITLPEALQRIPLATE
DIVRAPLAQTRLVKRVFGIDFLPHQTVHNTVRKGMQAHFHNARNAVSQAMR
WSLPYKXSQLDADCEAMDILGYLVQVRSQDQBNISDLSSSHILGQVREG"
BASE COUNT
537 a 629 c 534 g 501 t
ORIGIN
Query Match 76.9%; Score 1481.2; DB 10; Length 2201;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1503; Conservative 0; Mismatches 18; Indels 1; Gaps 1;
OY 402 AGGATGATCTGTGTAAGAAAGAGAGCGTGTGATGTTCTGGGTTCCAGGTCATCGTT 461
Db 316 AGATGATCTGTGTAAGAAAGAGAGCGTGTGATGTTCTGGGTTCCAGGTCATCGTT 375
OY 462 GTAGCTCTTCATCATATGTCGTCACAGACACCTTCCAGAGGAGAGATCCAG 521
Db 376 GTAGCTCTTCATCATATGTCGTCACAGACACCTTCCAGAGGAGAGATCCAG 435
OY 522 AGCCCGGTGATGCTGTGCTGTCTCTGCGGTCAAGATCTCTTTTGTGGACAG 581
Db 436 AGCCCGGTGATGCTGTGCTGTCTCTGCGGTCAAGATCTCTTTTGTGGACAG 495
OY 582 CTTTTCGGGACACACCGGATGTGTTCTACCTGATGAGACCTGCGCATGTGTGATG 641
Db 496 CTTTTCGGGACACACCGGATGTGTTCTACCTGATGAGACCTGCGCATGTGTGATG 555
OY 642 ACTTTCACGACGACGACCTGTGAAGCTGACATGCTGTGCGGAGATCTTCCGCTTC 701
Db 556 ACTTTCACGACGACGACCTGTGAAGCTGACATGCTGTGCGGAGATCTTCCGCTTC 615
OY 702 GTCCTTCGTGTGATGATGAGCGTCTTGTGATGCTATGAAACCGAGCCCGGAAACAG 761
Db 616 GTCCTTCGTGTGATGATGAGCGTCTTGTGATGCTATGAAACCGAGCCCGGAAACAG 675
OY 762 TCCAGACCTTTCACAGTGGAGCAAGCCGCGGCTGTGCTCAGCCCTGTGTGATCTTC 821
Db 676 TCCAGACCTTTCACAGTGGAGCAAGCCGCGGCTGTGCTCAGCCCTGTGTGATCTTC 735
OY 822 TTCCCTGCCACGAGATCAGTCCACCAAGCACTGCAAGCTGCTGCGGTGACAGCC 881
Db 736 TTCCCTGCCACGAGATCAGTCCACCAAGCACTGCAAGCTGCTGCGGTGACAGCC 795
OY 882 TTGATATGCTGAGAGAGCCCTGCGCTCTCAGGCTTCTGCTACTCAAGAGAGTGGCT 941

```

Db	796	TTTATATATGGTGGAGGAAGCCCTGCCGCTCTACGGCTTCGTGTACTCAAGAGAGTGGCT	855
OY	942	TTTCTCAGCCTGCAGGCCCTCTATCCACTACACGAGACCCCTTCCTCAACCTGCACGTC	1001
Db	856	TTTCTCAGCCCTGCAGGCCCTCTATCCACTACCTCAGAGACCCCTTCCTCAACCTGCACGTC	915
OY	1002	GTGCACCTGGTCCGAGAACCCCCGGGGCGTGTCCGATCCGGGGAGCAACACCATATGA	1061
Db	916	GTGCACCTGGTCCGAGAACCCCCGGGGCGTGTCCGATCCGGGGAGCAACACCATATGA	975
OY	1062	CTCATGTGTGACAGTCATATTTGTGCTAGGGACGATTTGGAAAGCATCAAGAGAGAAC	1121
Db	976	CTCTGTGGTTGACAGTCATATTTGTGCTAGGGACGATTTGGAAAGCATCAAGAGAGAAC	1035
OY	1122	CAGCCCTATTTATGCGATGAAATCATCATCGCAAAAGCCGAGGGGACATATGCTCAAGGCCATC	1181
Db	1036	CAGCCCTATTTATGCGATGAAATCATCATCGCAAAAGCCGAGGGGACATATGCTCAAGGCCATC	1095
OY	1182	CAAAACCCCTCCCTGAAGCTCTGCAGCAGCGCTACCTGTCTCCTAGAGATAGAGACCTGGTT	1241
Db	1096	CAAAACCCCTCCCTGAAGCTCTGCAGAGCGGTACCTGTTCCAGAGTATGAGAGACCTGGTT	1155
OY	1242	CGGGCACCCTCCGGCCGAGACGACCAAGACATATATAATTTGTGGGTTTGGATTTTGGCCC	1301
Db	1156	CGGGCACCCTCCGGCCGAGACGACCAAGACATATATAATTTGTGGGTTTGGATTTTGGCCC	1215
OY	1302	CACCTCCAAACATGGGTTTACATATGTCAACCCGCGCAAGGGCATGGGTACAGATCCCTC	1361
Db	1216	CACCTCCAAACATGGGTTTACATATGTCAACCCGCGCAAGGGCATGGGTACAGATCCCTC	1275
OY	1362	CATCTAAACGCCAGAGAACGCCCTCAAGCTCTCTCAGCGGTGGCGTTGGTCTTACCTTAC	1421
Db	1276	CATCTAAACGCCAGAGAACGCCCTCAAGCTCTCTCAGCGGTGGCGTTGGTCTTACCTTAC	1335
OY	1422	GAAAAGGTTTCCGACGCTCAAGATGCCCGGGTGGGGTATGGATTTCTGGGAATACCTC	1481
Db	1336	GAAAAGGTTTCCGACGCTCAAGATGCCCGGGTGGGGTATGGATTTCTGGGAATACCTC	1395
OY	1482	CAGGTCAGATCTCAACAAGAACCAAGGCAACCTGTCCCTGGATCTTCTGTCTCCTCCCAT	1541
Db	1396	CAGGTCAGATCTCAACAAGAACCAAGGCAACCTGTCCCTGGATCTTCTGTCTCCTCCCAT	1455
OY	1542	ATCTTGGGGGCAAGGCTTCCGAAAGGTTAAAGAGGTGCTGCACCCCTTGTTCCACG	1601
Db	1456	ATCTTGGGGGCAAGGCTTCCGAAAGGTTAAAGAGGTGCTGCACCCCTTGTTCCACG	1515
OY	1602	CTTAGTCACCTTAAACGACACAGAAAGCCTTAAGATATACCAAACTGAGTACCCTTCT	1661
Db	1516	CTTAGTCACCTTAAACGACACAGAAAGCCTTAAGATATACCAAACTGAGTACCCTTCT	1575
OY	1662	CCTCAGACCCCAAGCAGAGAGGGCTCTTTGTGTCTATATCATGTCTTACCTACATGAGCC	1721
Db	1576	CCTCAGACCCCAAGCAGAGAGGGCTCTTTGTGTCTATATCATGTCTTACCTACATGAGCC	1635
OY	1722	TAAAAAGCCAGAAACAGTATCTTCTCTGTAAGAAATACCTTAGGAACCTTAAGAGAGCC	1781
Db	1636	TAAAAAGCCAGAAACAGTATCTTCTCTGTAAGAAATACCTTAGGAACCTTAAGAGAGCC	1695
OY	1782	CTTTGACCTGTCAAGCAAGACTTCTTTTAACTTTGGCCTTCTTACCTGTGCATACCTTG	1841
Db	1696	CTTTGACCTGTCAAGCAAGACTTCTTTTAACTTTGGCCTTCTTACCTGTGCATACCTTG	1755
OY	1842	GAGACT-CGGTCTGGAGGCACTACGTGACACAGCAAAACGATCTGTGGAGTGTGTCTTA	1900
Db	1756	GAGACTACGGTCTGGAGGCTTACGTGACACAGCAAAACGATTCATGAGTGTGTCTTA	1815
OY	1901	AACCTCCCTGTACACATCTTTC	1922
Db	1816	AACCTCCCTGTACACATCTTTC	1837

AC130174/c LOCUS	232745 bp	DNA	linear	HTG 13-NOV-2002
DEFINITION	Rattus norvegicus clone CH230-14J11, WORKING DRAFT SEQUENCE, 3 unordered pieces.			
ACCESSION	AC130174			
VERSION	AC130174.5 GI:24635752			
KEYWORDS	HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTROP.			
SOURCE	Rattus norvegicus (Norway rat)			
ORGANISM	Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
REFERENCE	1 (bases 1 to 232745)			
AUTHORS	Muzny,D,Marie, Metzker,M,lee, Abramzon,S, Adams,C, Alder,J, Allen,C, Allen,H, Alshrooks,S, Amin,A, Angiano,D, Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Bernatmed,F, Bisswalo,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Censar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, d'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Ande,C, Dederich,D, Delgado,O, Denison,S, Deramo,C, Ding,Y, Dim,H, Diya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Gabregorgis,E, Geer,K, Gill,R, Grady,M, Guerra,M, Guerrero,W, Gunaratne,P, Haland,W, Hamil,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hugues,M, Hollins,B, Howells,S, Hulik,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jang,H, Johnson,B, Johnson,R, Jolivet,A, Kapathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowis,K, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorensunwa,L, Louiseged,H, Lozado,R,J, Lu,X, Ma,J, Maeswami,M, Mahndartine,M, Mamoud,M, Malloy,K, Mangum,A, Mangun,B, Mapua,P, Martin,K, Martin,R, Martinez,E, Mahlhley,S, McLeod,M,P, McNeill,T,Z, Meenen,E, Milosavljevic,A, Miner,G, Mijaj,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Mundasa,M, Murphy,M, Nair,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nwaekeluen,O, Okwuonu,G, Olarnpusagoon,A, Pal,S, Parks,K, Pasternak,S, Paul,H, Perez,A, Perez,L, Pfannkoch,C, Plopper,F, Poindexter,A, Popovic,D, Primus,E, Pu,L, Puzoz,M, Quiroz,J, Rachlin,E, Reeves,K, Rejzler,M,A, Relgh,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rives,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J, Sanders,M, Savery,G, Scherer,S, Scott,G, Shatsman,S, Shen,H, Shetty,J, Shvartsbeyn,A, Sisson,I, Sitter,C,D, Smajs,D, Steed,A, Sodergren,E, Song,X,-Z, Sorelle,R, Sosa,J, Steidle,M, Strong,R, Sutton,A, Svatek,A, Taboc,P, Taylor,C, Taylor,T, Thomas,N, Thomas,D, Tinley,A, Trefos,Z, Usmani,K, Valas,R, Vera,V, Villasana,D, Waldron,L, Walker,B, Wang,J, Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F, Williams,G, Willison,R, Wlczek,K,R, Wooden,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Dunn,D, von Niederhausen,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O, Weinstein,G, and Gibbs,R.A.			
TITLE	Direct Submission			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 232745)			
AUTHORS	Worley,K,C.			
TITLE	Direct Submission			
JOURNAL	Submitted (08-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA			
REFERENCE	3 (bases 1 to 232745)			
AUTHORS	Rat Genome Sequencing Consortium.			
TITLE	Direct Submission			

Db 175579 GGTCTTACCTTATACAAAGGTTTCTCAGCTTCAAGATGACTGTGTGACACTATGAT 175520
 Oy 1468 TGTGGATACCTCAGGTGATCTCAACAGAAAGGACACTGCTGATCTC 1527
 Db 175519 TTCTGGATACCTCAGGTGATCTCAACAGAAAGGACACTGCTGATCTC 175460
 Oy 1528 TGTCTCTCTCCATATCTTGGGGAGGTCTTCCAGAGAGTTAAGAGGTCTGTCTGAC 1587
 Db 175459 TGTCTCTCTCCATATCTTGGGGAGGTCTTCCAGAGAGTTAAGAGGTCTGTCTGAC 175400
 Oy 1588 CCTTGGTCCAGCTTATGATCACTTAACGACAGAGCT----TAAGTATACCA 1643
 Db 175399 CCTTGGTCCAGCTTATGATCACTTAACGACAGAGCTTTAAGTATGACCA 175340
 Oy 1644 AACTGAGTGGCTTCTCTCAGAGGCTTCTTGTGCTGATACATGAT 1703
 Db 175339 AACTGAGTGGCTTCTCTCAGAGGCTTCTTGTGCTGATACATGAT 175283
 Oy 1704 CTACCTTACCACTAGAGCTTAAAGGACAGAGATCTTCTGCTGAAATACCT 1763
 Db 175282 CTA---TACGGCCAAAGCTTAAAGGACAGAGATCTTCTGCTGAAATACCT 175226
 Oy 1764 AGGACCTTAAAGGACAGGCTTGAAGCTTGAAGCAAGA-----CTTCTTGAAGCTT 1816
 Db 175225 AGGACCTTAAAGGACAGGCTTGAAGCTTGAAGCAAGA-----CTTCTTGAAGCTT 175166
 Oy 1817 GGCCTTCTTACCTGATGATCTTGAAGCTTGAAGCAAGA-----CTTCTTGAAGCTT 1875
 Db 175165 AGCCTTCTTACCTGATGATCTTGAAGCTTGAAGCAAGA-----CTTCTTGAAGCTT 175106
 Oy 1876 ACAGCATCTGTGAGGTGTCTGTAACCTCTG 1910
 Db 175105 GCAGCATCTGTGAGGTGTCTGTAACCTCTG 175071

RESULT 5
 AC110314/c 262679 bp DNA linear HTG 20-NOV-2002
 LOCUS Rattus norvegicus clone CH230-198124, WORKING DRAFT SEQUENCE, 5
 DEFINITION unorderd pieces.
 ACCESSION AC110314
 VERSION AC110314.5 GI:25139563
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 262679)
 Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amlin, A., Angiano, D.,
 Anulobechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Barnstead, F.,
 Batswala, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M., L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 George, E., Geier, K., Gill, R., Grady, M., Guerra, W., Guera, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Hulik, S., Hume, J., Idelbird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowals, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 JOURNAL

REFERENCE
 AUTHORS
 JOURNAL
 COMMENT

Lorenshewa, L., Louised, H., Lozano, R. J., Lu, X., Ma, J.,
 Maheshwari, M., Mahdarte, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapur, P., Martin, K., Martin, R., Martinez, E.,
 Mahoney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minis, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Mundasa, N., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwokelam, O., Okunolu, G., Olajun, A., Pal, S., Parks, K.,
 Paternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Poldner, A., Popovic, D., Prins, E., Pu, L.,
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rites, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Slason, I., Slicker, C. D., Smaj, D.,
 Sneed, A., Sodergren, E., Song, X. Z., Sorlie, R., Sosa, J.,
 Stead, M., Strong, R., Sutton, A., Swale, A., Tabot, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wiczyski, R., Wooden, H., Wrobley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
 Weinstock, G. and Gibbs, R. A.
 Direct Submission
 Unpublished
 2 (bases 1 to 262679)
 Wrobley, K. C.
 Direct Submission
 Submitted (11-FEB-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 262679)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 20, 2002 this sequence version replaced gi:23101413.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GOGM
 Center clone name: CH230-198124
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 234127 bases at least Q40
 Consensus quality: 237681 bases at least Q30
 Consensus quality: 239914 bases at least Q20
 Estimated insert size: 237579; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_difc_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

OY	1093	AGCATTGGAAACATCAAGAGAAACACAGCCATTATGCGCATGAAGATCATCGCA	1152		
Db	63977	AGAAATTGAAAAATATCATAAAGAGAAAGACCAACCCCTATTATGCGCATGAAGATCATCGCA	63911		
OY	1153	AAAGCCAGGTGACATAGTCAAGGCCATCCAAACCCCTCCCTGAAAGCTGTGCACAGCGCT	1212		
Db	63917	AAAGCCAGGTGACATAGTCAAGGCCATCCAAACCCCTACTGTAAAGTTCTGCACAGCGCT	63855		
OY	1213	ACCGTTCCTGAGGTATAGAGACCTGGTTGGGGCACCCTCGGCCACAGAGCACCAAGCTAT	1272		
Db	63857	ACCGTTCCTGAGGTATAGAGACCTGGTTGGGGCACCCTCTGAGTCAAGACTTCCAGACTGT	63798		
OY	1273	ATAAATTTGGGGGTGGATTTTGTGGCCACCCCTCCAAACATGGTTTACATATGTACACC	1332		
Db	63797	ATAAATTTGGGGGTGGATTTTGTGGCCACCCCTCTGATCATGTCCAAACCTGGTTTCAACATGTACACC	63738		
OY	1333	GCGGCAAGGCAATGGGTGCAGCATTGCTTCATACTAAGCCGCAAGAAAGCCCTTCACGTCT	1392		
Db	63737	GTGGCAAGGCGCATGGGTGCAGCATTGCTTCATACTAAGCCGCAAGAAAGCCCTTCACGTCT	63678		
OY	1393	CTCAGGCGTGGCGGTGGTGGCTTACCTTACGAAAGGTTGCCAGTCCCAAGATGCTCGG	1452		
Db	63677	CCACAGCATGGCGGTGGTGGCTTACCTTACCTTATACAAAGGTTCTCAGCTTCAAGATGCTGTG	63618		
OY	1453	GTGAGGCTATGATTTGCTGGGATACCTCCAGCTCAGATCTCAACAAAGCAAGGCAAC	1512		
Db	63617	GTGAAAGCTATGATTTCTGGGATACCTCCAGCTCAGATCTCAACAAAGCAAGGCAAC	63558		
OY	1513	TGTGCTCGTGAATCTTCTGTCTCTCCCATATCTTGGGGCAGGTCTTCCGAAAGTTAAG	1572		
Db	63557	TGTGCTCGTGAATCTTCTGTCTCTCCCATATCTTGGGGCAGGTCTTCCGAAAGTTAAG	63498		
OY	1573	GAGGTCTGTGTGCACCCCTGGTGTCCAGCCTTAGTACCATTTAAAGCGCAGAAAGCT--	1630		
Db	63497	GAGGTCTGTGTGCACCCCTGGTGTCCAGCCTTAGTACCATTTAAAGCGCAGAAAGCTTA	63438		
OY	1631	--TAAAGTATTAACCAACTGATGAGCCCTTCTCTCAGCCCCCAAGAGGGGTCTTGG	1688		
Db	63437	AGTAAAGTATTAACCAACTGAGCACCATTATCTCAGTCCCAAGAGGAGTCTCT--	63379		
OY	1689	TGTCTATATCACTAGTGTACCCCTACACTGAGCCTAAAGAGCCAAAGAAACAGTATCTTCT	1748		
Db	63378	--TTTATATCACTAGGTCTA--TACGGCCAAAGCCTTAAAGGCCAAAGAAATATCTTCT	63324		
OY	1749	GCTCTGAATAATAC	1761		
Db	63323	GCTCTGAACATTC	63311		
RESULT 6	AC010547	183228 bp	DNA	linear	PRI 29-SEP-2001
LOCUS	AC010547				
DEFINITION	Homo sapiens chromosome 16 clone RP11-510M2, complete sequence.				
ACCESSION	AC010547				
VERSION	AC010547.9				
SOURCE	GTG				
KEYWORDS	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 183228)				
TITLE	DOE Joint Genome Institute.				
JOURNAL	Sequencing of Human Chromosome 16				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 183228)				
TITLE	DOE Joint Genome Institute.				
JOURNAL	Direct Submission				
REFERENCE	Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA				
AUTHORS	3 (bases 1 to 183228)				
TITLE	DOE Joint Genome Institute.				
JOURNAL	Direct Submission				

DEFINITION Homo sapiens L-selectin ligand sulfotransferase mRNA, complete cds.
 ACCESSION AF149783
 VERSION AF149783.1 GI:13897503
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1333)
 AUTHORS Yeh, J.C., Haraoka, N., Petryniak, B., Nakayama, J., Elles, L.G.,
 Rebuta, D., Hindsgraul, O., March, J.D., Lowe, J.B. and Fukuda, M.
 Novel sulfated lymphocyte homing receptors and their control by a
 Core1 extension beta 1,3-N-acetylglucosaminyltransferase
 CELL 105 (7), 957-969 (2001)
 JOURNAL MEDLINE
 PUBMED 21332592
 REFERENCE 2 (bases 1 to 1333)
 AUTHORS Haraoka, N. and Fukuda, M.
 TITLE Direct Submission
 JOURNAL Submitted (10-MAY-1999) Glycobiology, The Burnham Institute, 10901
 North Torrey Pines Road, La Jolla, CA 92037, USA
 FEATURES
 source
 location/Qualifiers
 1..1333
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 93..1253
 /note="LSSST"
 /codon_start=1
 /product="L-selectin ligand sulfotransferase"
 /protein_id="AAK48417.1"
 /db_xref="GI:13897504"
 /translation="MLPKMKLLFLVSOMALALFFHWYSHNISLSMKAOERH
 VILSSWRSSSPYGLFQHPDYFLMEPMHYMPKFOSTAMNLMAYDILRAY
 LCDMSVDATMEPRPROSLFQWENRALSAPACITIPDEILIPRACHLCSQP
 FEVEKACRSYSHVLEKVERFENLQSLYPLKIDSLNLVHLVRAVFRSLERK
 GDLMDIRISYVGOHEOKLREKEDPYVMYICOSLEIKYIOLPLPALDERLYLVRY
 EDLARAVAQTSRYMEFVGLFELPHLOTVHNITRGGMGHAFHTNARDLANSQAM
 RMSLPYKRVRLQKACGDAMNLLGVRHVRSHQEQRNLLDLSTWYVEQIH"
 BASE COUNT 286 a 393 c 336 g 318 t
 ORIGIN
 Query Match 39.0%; Score 751.2; DB 9; Length 1333;
 Best Local Similarity 76.0%; Pred. No. 1.7e-188;
 Matches 968; Conservative 0; Mismatches 298; Indels 8; Gaps 3;

757 AACAGTCCAGCCTCTTCCAGTGGAGCAAGCCGGCCCTGCTGCTCAGCCGCTGTGTG 816
 448 GACAGTCCAGCCTCTTCCAGTGGAGCAAGCCGGCCCTGCTGCTGCTGCTGCTGCTG 507
 817 ACTTCTTCCCGCCAGAGATGACATCCAGCAACATGCACTGCTGCTGCTGCTGCTGCTG 876
 508 ACATCATCCCAAGATGAATCATCCCGCGGCTCACTGAGGCTCTCTGTGCAATGAC 567
 877 AGCCCTTGATATGATGAGAGAGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 936
 568 AGCCCTTGATATGATGAGAGAGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 627
 937 TGGCTTTTCTCAGCTGAGGCGCTCTATCCACTGACAGGAGCCCTTCCCTCAACCTGC 996
 628 TGGCTTTTCTCAGCTGAGGCGCTCTATCCACTGACAGGAGCCCTTCCCTCAACCTGC 687
 997 AGCTGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1056
 688 ATATGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 747
 1057 TAGAATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1116
 748 GAGATCTCATGATGACAGTGCATGTGATGAGGCGACATGACCAAACTCAAGAGG 807
 1117 AAGACAGCCCTATTATCCATGAAGATCATCTGCAAAAGCAGAGTGACATGATCAAGG 1176
 808 AGGACCAACCTACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 867
 1177 CCATCCAAACCTCCCTGAGACCTGTCAGACAGCCCTGCTGCTGCTGCTGCTGCTGCTG 1236
 868 CCATCCAAACCTCCCTGAGACCTGTCAGACAGCCCTGCTGCTGCTGCTGCTGCTGCTG 927
 1237 TGGTTCGGGACCCCTGGCCCAAGACCAACATATATATTTGTGGGATGATTTT 1296
 928 TGGTTCGGGACCCCTGGCCCAAGACCAACATATATATTTGTGGGATGATTTT 987
 1297 TGGCCCAACCTCCCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1356
 988 TGCCCAACCTCCCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1047
 1357 CCTTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1416
 1048 CTTTCCAAACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1107
 1417 CTTAGCAAAAGGTTTCCAGCTGAAGTGGCTGGGAGGATGATGATGATGATGATGATGATG 1476
 1108 CTTAGCAAAAGGTTTCCAGCTGAAGTGGCTGGGAGGATGATGATGATGATGATGATGATG 1167
 1477 ACCTCAGAGTGAATCTTCACAAAGCAAGCAACCTGCTGCTGCTGCTGCTGCTGCTGCTG 1536
 1168 ACCGCAAGTGAATCTTCACAAAGCAAGCAACCTGCTGCTGCTGCTGCTGCTGCTGCTG 1227
 1537 CCATATCTTGGGAGGCTGCTTCCGAAAGGTTAAGAGGCTGCTGCTGCTGCTGCTGCTG 1596
 1228 GGAATGCTTCCGAAATTCACTAAGAGGTTGAAGAGGCTTGGCTG-CCACTGGTG 1285
 1597 CCAGCTTATGAC 1610
 1286 TCAAGCTCAGTAC 1299

RESULT 9
 AF280088 1992 bp mRNA linear PRI 20-FEB-2001
 LOCUS Homo sapiens L-selectin ligand sulfotransferase GST-3 mRNA,
 complete cds.
 ACCESSION AF280088
 VERSION AF280088.1 GI:12060807
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 1992)
 AUTHORS Hemmerich, S., Lee, J.-K., Bhakta, S., Bistrup, A., Ruddle, N.R. and Rosen, S.D.
 TITLE Chromosomal localization and genomic organization for the galactose/N-acetylgalactosamine/N-acetylglucosamine 6-O-sulfotransferase gene family
 JOURNAL Glycobiology 11 (1), 75-87 (2001)
 MEDLINE 21096027
 PUBMED 11181564
 REFERENCE 2 (bases 1 to 1992)
 AUTHORS Hemmerich, S., Lee, J.-K., Bistrup, A., Ruddle, N.R. and Rosen, S.D.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience, 3401 Hillview Ave., Palo Alto, CA 94304, USA
 FEATURES
 source 1.1992
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="16"
 /map="16q23.1-q23.2"
 /cell_type="high endothelial"
 /tissue_type="Conail"
 /gene="GST3"
 /gene="GST3"
 /note="similar to Homo sapiens sequence encoded by GenBank accession number AF131235"
 /codon_start=1
 /product="L-selectin ligand sulfotransferase GST-3"
 /protein_id="AAG48246.1"
 /db_xref="GI:12060808"
 /translation="MLPKKMLLEFVVSOMALIAFPMVSHNLSLSMKAPENH
 VLVLSMRGSSFFVQULBQGHVDVITMEPRANHWTFKOSTAMLMHVRDLIRAF
 LCDMVFDAIMEPRROSILFOWENSRLCSAPACDILPDEILIPRARCILCSOP
 FEVVKACSYSHVVLAKREFNLSLPLADPSLNLIVHLVDRPRAVFRSRETK
 GDLMDISRLVMOHOKLEKEDQPYVMQVVICOSQLEITKIQSLPKALOERYLVRY
 EDLAPAPVQATSRMEFVGLFELPLHLOTGVNHTGKMGDHAFTNADLVNSQAW
 RMSLPYKVSRLQKACGDAMNLGVRHVSDEQNRLLDLSTMTVPQIH"
 BASE COUNT 482 a 540 c 489 g 481 t
 ORIGIN
 Query Match 38.6%; Score 743.2; DB 9; Length 1992;
 Best Local Similarity 76.5%; Pred. No. 2.3e-186;
 Matches 938; Conservative 0; Mismatches 283; Indels 5; Gaps 2;

QY 745 CAGCCCCCGGAAGACGTCACGCTCTTCCAGTGGAGCAAGCCGGCCCTGTGCTCAG 804
 Db 465 CTGGTCCCGGAGACAGTCACGCTCTTCCAGTGGAGCAAGCCGGCCCTGTGCTCAG 524
 QY 805 CCGCTGTGTGCTCTTCCCTCCAGAGATCAGTCACCAAGCAAGCTGCAACTGC 864
 Db 525 CACCTGCTGTGATCATCATCCCAAGATGAATATCCCGGGCTCATGTGAGCTCC 584
 QY 865 TCTGCGGTGAGCAGCCCTTGTGATGTGGAGAGAGCCCTCCCTCAGCGCTTGTG 924
 Db 585 TGTGAGTCAAGACGCCCTTGTGAGTGTGGAGAGAGCCCTCCCTCAGCGAGTGG 644
 QY 925 TACTCAGAGAGTGTGCTTCTCAGCTGAGGCGCTTATCCATCTACAGCACTT 984
 Db 645 TGTCTCAGAGAGTGTGCTTCTCAGCTGAGGCGCTTATCCATCTACAGCACTT 704
 QY 985 CCCTCAACCTGACAGCTGTGACAGCTGTGAGAGAGCCCGGGCGGTCCGATCCGG 1044
 Db 705 CCCTCAACCTGACATATGTGACAGCTGTGAGAGAGCCCGGGCGGTCCGATCCGG 764
 QY 1045 AGCAGACCAACATAGAACTGATGATGATGATGATGATGATGATGATGATGATG 1104
 Db 765 AACGACAAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 824
 QY 1105 CGATCAG 1164
 Db 825 AACTCAAG 884
 QY 1165 AACTCAAG 1224
 Db 885 AACTCAAG 944
 QY 1225 GGTATGAG 1284
 Db 945 GGTATGAG 1004
 QY 1285 GGTATGAG 1344
 Db 1005 GGTATGAG 1064
 QY 1345 TGGGTGAG 1404
 Db 1065 TGGGTGAG 1124
 QY 1405 GTTGTGCTTACCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1464
 Db 1125 GGTGTGCTTACCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1184
 QY 1465 ATTTGTGCTTACCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1524
 Db 1185 ATTTGTGCTTACCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1244
 QY 1525 TTCTGTGCTTACCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1584
 Db 1245 TTCTGTGCTTACCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1304
 QY 1585 CACCCCTGTGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1640
 Db 1305 --CACTGTGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1328
 RESULT 10
 AK026635
 LOCUS AK026635 2011 bp mRNA linear PRI 29-SEP-2000
 DEFINITION Homo sapiens cDNA: FLJ22982 fls, clone KRL1454, highly similar to AF131235 Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA.
 ACCESSION AK026635
 VERSION AK026635.1 GI:10439531
 KEYWORDS oligo capping; fls (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REMARK
COMMENT

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgabs-rr@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amgdbcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Huiyik, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
 Series: IRM Plate: 50 Row: a Column: 7
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 5031734.
 Location/Qualifiers

FEATURES
source

1..2037
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:34456 IMAGE:5179826"
 /tissue_type="Brain, Lung, Testis, adult, pooled whole"
 /clone_id="NIH_MGC_115"
 /lab_host="DH10B"
 /note="vector: pCMV-Sport6"
 /size="1238"
 /codon_start=1
 /product="Similar to carbohydrate (N-acetyl)glucosamine 6-O sulfotransferase 4"
 /protein_id="AAH35282.1"
 /db_xref="gi:23273965"

CDS

126..1238
 /product="Similar to carbohydrate (N-acetyl)glucosamine 6-O sulfotransferase 4"
 /protein_id="AAH35282.1"
 /db_xref="gi:23273965"
 /translation="MALALFPHMYSNHSISLAKAPRHHVLYSSKSSSTVGO
 LFGHPDVFYLEPAMHVMWTKOSTAMWLAHVBDLRAVFLCMSVYDAYMEPR
 ROSSLFOWENRSLALSAACDIIPDEIIPRAHCRILSQDFEVEKCRSYSHVL
 KEVFENLYSLPLKADPSLNIHVLVDPRAVRSRRTGDLMDIRIVMGQHEO
 KLKEDOPYSYVMVQCSOLEYKTIOSPKALORRYLLVREDLARAVALTSRYE
 FVGLFELHLOTWVHNITRGKMGDHAFTNADALNSQANRMLPYEKVSRLLKAC
 GDAMLLFHLTRHRSQEQRNLLDLSTVTPYEQIH"

BASE COUNT
ORIGIN

Query Match 38.6%; Score 743.2; DB 9; Length 2037;
 Best Local Similarity 76.5%; Pred. No. 2,3e-186;
 Matches 938; Conservative 0; Mismatches 283; Indels 5; Gaps 2;

388 TCTCTCTTCGCGAGATGATGCTTGAGAAAGGAGGCTGTGATGTTCTGGGTT 447
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 61 TCTTCACCTTCAGCAATGCTACTGCTAAAAAATGAAGCTCGTGTTCGTGTT 120
 448 CCCAGGTATCTGTAGCTCTCTCATATGCTCCGACAGACACCTTCCAG- 506
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 121 CCCAATGCGCATCTGTGCTATCTTCACATGACCAACCAATCAGCTCCCTGT 180
 507 --AGGAGAGATCCAGAGAGCCGTCATGCTGTGCTGCTCTCTCCGCGGTGAGAT 564
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 181 CATAGAGGACAGCCCGAGCGCATGACGCTGTGCTGTCTCTCTGCGCTCTGCT 240
 565 CCTCTTTGTGGAGAGCTTTTCGGGACAGCACCGGATGTTCTTACATGATGAGCCTG 624
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 241 CTTCTTTGTGGGAGCTTTTGGGACAGCACCGATTTTCTTACATGATGAGCCCG 300
 625 CCTGCATGTGTGATGATCTTACAGACAGACCTGGAAGCTGACATGCTGTGC 684
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 301 CCTGCACGTGTGATGATGATCTTCAAGCAGACCGCGTGAATGCTGACATGCTGTGC 360
 685 GGGATCTTCTGCGTCCGCTCTCTGTGTGATGATGAGGCTTTGATGCTCATGAGAAC 744

||||| 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 361 GGGATCTGATGAGGCGCGCTCTTCTGTGAGATGAGCTTTGATGCTCATGAGAAC 420
 745 CAGGCCCCGGGAAGAGTCCAGCTCTTCAGGGGAGAAAGCCGGCCGTGTGAG 804
 421 CTGTGCCCCGAGACAGTCCAGCTCTTTCAGGGGAGAAAGCCGGCCGTGTGAG 480
 805 CCGCTGTGTGATGATCTTCTCCGTCAGAGATGAGTCCAGCAGCAGTCCAGCTGC 864
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 481 CACCTGCTGTGATGATCATCCAGAGATGAGAAATATCCCGGGGCTCACTGAGGCTCC 540
 865 TCTGCGGTGAGAGCCCTTTGATATGTTGAGAGAGCCCTGCGCTCTCAGGCTTGTG 924
 541 TGTGAGTCAAGAGCCCTTTGAGTGTGAGAGAGCCCTGCGCTCTCAGCAGCTGTG 600
 925 TACTCAAGAGGAGCTTTTCAGGCTGAGAGCCCTTATGACACTACAGAGCCCTT 984
 601 TGTCTAAGAGAGTGTGCTTCTTCAACCTCAGTCCCTTACCGGCTGTGAAAGACCTT 660
 985 CCGTCAACCTGACAGTGTGACAGCTGTGTCGAGAGCCCGGGCCGTGTCCGATCCGGG 1044
 661 CCGTCAACCTGACATATGTCAGCTGTGTCGAGAGCCCGGGCCGTGTCCGATCCGGG 720
 1045 AGCAGCAGCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1104
 721 AACGCAAGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
 1105 CGATCAAG 1164
 781 AACTCAAG 840
 1165 AACTCAAG 1224
 841 AACTCAAG 900
 1225 GGTATGAG 1284
 901 GGTATGAG 960
 1285 GGTATGAG 1344
 961 GGTATGAG 1020
 1345 TGGGTGAG 1404
 1021 TGGGTGAG 1080
 1405 GTTGTGAG 1464
 1081 GGTGTGAG 1140
 1141 ATTGTGAG 1200
 1525 TTTGTGAG 1584
 1201 TTTGTGAG 1260
 1585 CACCCCTGAG 1640
 1261 --CCACCTGAG 1284

RESULT 12
 BD127258
 LOCUS BD127258 1979 bp DNA linear PAT 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD127258
 VERSION BD127258.1 GI:23222203
 KEYWORDS JP 200201737-A/2689.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1979)
Ota, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
Makamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and
Koga, H.,
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002017375-A 2689 22-JAN-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002017375-A/2689
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253172
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
PI ISHII,
PI YURI KAWAI, AI MAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUJI OTSUKI, HISASHI KOGA
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/PC
10,
PC C12P21/02, C1201/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
(69). (1226).
FT CDS Location/Qualifiers
1..1979
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT 451 a 542 c 474 g 512 t
ORIGIN

Query Match 38.5% Score 741.6; DB 6; Length 1979;
Best Local Similarity 76.4%; Pred. NO. 6.2e-186;
Matches 937; Conservative 0; Mismatches 284; Indels 5; Gaps 2;

QY 388 TCTCTCTTCCGAGATGATGCTTTGAAGAAAGGAGGCTGATGTTCTGGGT 447
DB 52 TCTTCCACTTCAGACAAATGCTACTGCTTAAAGAAAGGAGCTGCTGTTTGGTTT 111
QY 448 CCCAGGTCATGCTGTTAGCTCTCTTCAATCATATGCTCCACAGACACCTTCCAG- 506
DB 112 CCCAGATGGCCATCTTGCTCTATCTTCCATGACGACCAACATCATGCTCCCTGT 171
QY 507 --AGGGAGAGTCCAGAGGCGGCTGATGCTGCGGCTGCTTCTGGGGCAGAT 564
DB 172 CTATGAAGGCAAGCCGAGGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 231
QY 565 CCTCTTTTGGGACACCTTTTGGGAGACACCGGATGTTTACCTGATGAGGCTG 624
DB 232 CTCTTTTGGGGGACCTTTTGGGAGACACCGGATGTTTACCTGATGAGGCTG 291
QY 625 CCTGGCATGCTGATGATCTTTCACAGACAGCCCTGAGACCTGACATGGCTGTC 684
DB 292 CCTGGCAGCTGATGATGATCTTTCACAGACAGCCCTGAGACCTGACATGGCTGTC 351
QY 685 GGGATCTTGGCGTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 744
DB 352 GGGATCTTGGCGTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 411
QY 745 CAGGCCCCCGAAGACATCCAGCTTTCAGTGGAGGAGGAGGCGGCTGCTGCTG 804
DB 412 CTGTGCGCGGAGACATCCAGCTTTCAGTGGAGGAGGAGGCGGCTGCTGCTG 471
QY 805 CGCGCTGTGTGACTTCTTCCCTGCGCAGAGATCAGCTCAGCCAGCTCAGAGTGC 864
DB 472 CACCTGCTGTGATCATCTCCACAGAGATCAATCCCGGGCTCAGCTCAGAGTGC 531
QY 865 TCTGGGTCAGAGCCCTTGTATGATGATGAGAGGAGGCTGCTGCTGCTGCTG 924
DB 532 TGTGCACTCAACGCCCTTGTAGGAGTGTGAGAGGCTGCTGCTGCTGCTGCTG 591

QY 925 TACTCAAGAGAGTGCCTTTTCTCAGCCTGCAGGCCCTCTATCATCTACAGGACCTT 984
DB 592 TGTCTCAAGAGAGTGCCTTTTCTCAGCCTGCAGGCCCTCTATCATCTACAGGACCTT 651
QY 985 CCTTCAAGCTTGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1044
DB 652 CCTTCAAGCTTGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 711
QY 1045 AGCAGCACCACATGTAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1104
DB 712 AAGCAGCACCACATGTAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 771
QY 1105 CGATCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1164
DB 772 AACTCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 831
QY 1165 ACATAGTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1224
DB 832 AGATCTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 891
QY 1225 GGTATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1284
DB 892 GGTATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 951
QY 1285 GGTATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1344
DB 952 GATTGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1011
QY 1345 TGGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1404
DB 1012 TGGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1071
QY 1405 GTTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1464
DB 1072 GCTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1131
QY 1465 ATTGTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1524
DB 1132 ATTGTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1191
QY 1525 TTTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1584
DB 1192 TTTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1251
QY 1585 CACCCCTTGTGCTGAGGCTTGTAC 1610
DB 1252 --CCACTGTGTGAGGCTTGTAC 1275

RESULT 13
AK074746
LOCUS 1979 bp mRNA linear PRI 03-SEP-2002
DEFINITION Homo sapiens cDNA FLJ90265 f1s, clone NT2RM4001325, weakly similar
TO CHONDROITIN 6-SULFOTRANSFERASE (BC 2.8.2.17).
ACCESSION AK074746
VERSION AK074746.1 GI:22760388
KEYWORDS oligo capping; f1s (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS
1 Isogai, T., Ota, T., Nishikawa, T., Hayashi, K., Otsuki, T.,
Sugiyama, T., Suzuki, Y., Nagai, K., Sugano, S., Ishii, S.,
Kawai, H., Y., Saito, K., Yamamoto, J., Makamatsu, A., Nakamura, Y.,
Kojima, S., Nagahara, K., Masuno, Y., Ota, T., Okano, K., Yoshikawa, Y.,
Aotsuka, S., Sasaki, N., Hattori, A., Okumura, K., Iwayanagi, T. and
Nishimura, K.
TITLE NEDD human cDNA sequencing project
JOURNAL unpublished
REFERENCE 2 (bases 1 to 1979)
AUTHORS Isogai, T. and Otsuki, T.
TITLE Direct Submission

JOURNAL

Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction: Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.).

FEATURES

source

1. 1979
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RM4001325"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/note="cloning vector: pME185FL3-mRNA from uninduced NT2 neuronal precursor cells"
69. 1229
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC11177.1"
/db_xref="GI:22760389"
/translation="MLPKMKLLFLVSOMALALFHHYSHNISLSMKAPERMHVLVSSWRSSSEFVQLEQHPDVFYLMPPHVMWTFKQSTAMMLHMAVRDLIRAVFLDMVDFAYMERPRROSLFQWENSRALCSAPCDLIPDEIIPRAVRLCSQPFVEVEKACRSYSHVLLKEVFNLSLYLLKDPSELIHVLHVRPFRSLREKTKGLMDINRIVMGHOKLEKEDPYVYVICOSELEIKTIOSLPKALOEYLYLVKEDLAPVAVNOTSRMYEYVGLFEFLPHLQTYHNTTRKGGDIAPFHNAIDALNVSQAMRSLYEKYSRLQKACGDMNLLGRHNSBOQRNLLDLSTWTPQIH"

CDS

BASE COUNT

451 a 542 c 474 g 512 t

ORIGIN

Query Match 38.5%; Score 741.6; DB 9; Length 1979;
Best Local Similarity 76.4%; Pred. No. 6.2e-186;
Matches 937; Conservative 0; Mismatches 284; Indels 5; Gaps 2;

388 TCTCTCTCTCCGAGATGATGCTGTTGAAGAGGAGGCTGTGATGTTCCCTGGTT 447
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
52 TCTTCACATTCAGACAAATGCTACTGCTTAAAAAATGAAGCTCGCTGTTCTGTTT 111
111 111 111 111 111 111 111 111 111 111 111 111 111 111
448 CCCAGGTATCGTTGAGCTCTCTCATATGCTCCGTCCACAGACCTTTCCAG- 506
111 111 111 111 111 111 111 111 111 111 111 111 111 111
112 CCCAGATGGCCATCTTGCTCTATCTTCCATATGACGACCAACATCAGCTCCCTGT 171
111 111 111 111 111 111 111 111 111 111 111 111 111 111
507 --AGGAGAGATCCAGAGAGCCCGTGCATGTGCTGCTGCTCTGCGGTCAGAT 564
111 111 111 111 111 111 111 111 111 111 111 111 111 111
172 CATTGAAGGACAGCCCGGACGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 231
111 111 111 111 111 111 111 111 111 111 111 111 111 111
565 CCTCTTTTGTGGAGACTTTTGGGACACCCGAGTGTCTTACATGATGAGCCGTG 624
111 111 111 111 111 111 111 111 111 111 111 111 111 111
232 CTTCTTTTGTGGGAGCTTTTGGGACACCCAGATGTTTCTACATGATGAGCCCG 291
111 111 111 111 111 111 111 111 111 111 111 111 111 111
625 CCTGGATGTGTGATGATTTTACACAGACAGACCTGGAAGCTGCATATGCTGTGC 684
111 111 111 111 111 111 111 111 111 111 111 111 111 111
292 CCTGGACGTGTGATGATGATTTTACACAGACAGACCTGGAAGCTGCATATGCTGTGC 351
111 111 111 111 111 111 111 111 111 111 111 111 111 111
685 GGGATCTTCTGCTTCTGCTCTGATGATGATGATGATGATGATGATGATGATGATG 744
111 111 111 111 111 111 111 111 111 111 111 111 111 111
352 GGGATCTTCTGCTTCTGCTCTGATGATGATGATGATGATGATGATGATGATGATG 411
111 111 111 111 111 111 111 111 111 111 111 111 111 111
745 CAGGCCCCCGAAGACAGTCTTCTTCCAGTGGAGCAAGCCGCGCTGTGCTCAG 804
111 111 111 111 111 111 111 111 111 111 111 111 111 111
412 CTGGTCCCGGAGAGAGTCCAGCTTTTCAAGTGGAGCAAGCCGCGCTGTGCTCAG 471
111 111 111 111 111 111 111 111 111 111 111 111 111 111
805 CCGCTGTGTGATCTTCTTCCGCGCAGAGATCAGCTCACCAGACATGCAAGCTGC 864
111 111 111 111 111 111 111 111 111 111 111 111 111 111
472 CACTGTGCTGTGATCATCATCAAGATGAATATCATCCCGGCGCTCATGCAAGCTGC 531
111 111 111 111 111 111 111 111 111 111 111 111 111 111
865 TCTGGGTGACAGACCCCTTGTATGTGTGAGAGAGCCCTGCGCTCTCAGCGCTTGTG 924
111 111 111 111 111 111 111 111 111 111 111 111 111 111

111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 532 TGTGCAAGTCAAGACCCCTTGTGAGTGTGAGAGAGCCCTGCTCTACAGCAAGTGG 591
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Qy 925 TACTCAAGAGAGGCTTTTTCAGGCTGTCAGAGCCCTTATTCATCTACAGCAAGCTT 984
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 592 TGTCTCAAGAGAGTGGCTTCTTCAACCTGCTACCTGCTTACCCGCTGGAAGACCTT 651
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Qy 985 CCTCAACCTGCAAGCTGTCAGCTGTCAGAGACCCCGGCGCTGTCCGATCCGCGG 1044
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 652 CCTCAACCTGCAATATGTCAGCTGTCAGAGACCCCGGCGCTGTCCGATCCGCGG 711
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Qy 1045 AGCAGACCAACATATGATCATATGTTGAAGATCATATGTTGTAAGGACGATTTGGAA 1104
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 712 AACGACAAAGAGAGATCATATGATGAACATATGATGATGATGATGATGATGATGAT 771
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Qy 1105 CGATCAAG 1164
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 772 AACTCAAG 831
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Qy 1165 ACATAGTCAAG 1224
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 832 AGATCTACAAAGACATCCAGTCTCTCCCAAGGCGCTGCAAGAGAGAGAGAGAGAG 891
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Qy 1225 GGTATGAG 1284
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 892 GGTATGAG 951
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Qy 1285 GGTATGAG 1344
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 952 GATGGAATCTTGTCCCAATCTTCAACAGCTGGGTGATCAATCAACCCAGAGAGAGCA 1011
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Qy 1345 TGGGTAGAGATGCTCTTCACTACTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1404
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 1012 TGGGTAG 1071
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Qy 1405 GTTGTGCTTACCTTACGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1464
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 1072 GCTGTGCTTGTGCTTGTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1131
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Qy 1465 ATTGTGAGAGATCTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 1524
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 1132 ATTGTGAGAGATCTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 1191
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Qy 1525 TTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTG 1584
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 1192 TTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTG 1251
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Qy 1585 CACCCCTGTTCCAGCCTTATGTCAC 1610
111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 1252 --CCACTGTGTACACCTTCAAGTCAAC 1275
111 111 111 111 111 111 111 111 111 111 111 111 111 111

RESULT 14
AR203335 AR203335 2032 bp DNA linear PAT 20-JUN-2002
LOCUS AR203335
DEFINITION Sequence 2 from patent US 6365365.
ACCESSION AR203335
VERSION AR203335.1 GI:21499698
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2032)
Bistup,A., Rosen,S.D., Tangemann,K. and Hemmerich,S.
TITLE Method of determining whether an agent modulates glycosyl
sulfotransferase-3
JOURNAL Patent: US 6365365-A 2 02-APR-2002;
FEATURES Location/Qualifiers
source 1..2032
BASE COUNT 468 a 569 c 490 g 505 t
ORIGIN

Db 345 CTTCTTTTGGGGGACGTTTGGGACAGCCAGATGTTTCTACCTGATGAGCCCG 404
 QY 625 CCTGCATGTGTGATGACTTTCACAGCAGACAGCTGGAGCTGCACATGGCTGTGC 684
 Db 405 CCTGGACAGGTGGATGACCTTCAAGCAGACAGCCGTGATGCTGACATGGCTGTGC 464
 QY 685 GGGATCTTCCGCTTCCGCTTCTGCTGATGACATGAGCGCTTGTGATGCTACATGACC 744
 Db 465 GGGATCTGATGAGCGGCTTCTTCTTGTGACATGAGCGCTTGTGATGCTACATGACC 524
 QY 745 CAGGCCCCGGGAAACAGCTCCAGGCTTCCAGTGGAGCAAAAGCCGGCCCTGTGCTAG 804
 Db 525 CTGTGCCCCGGGAAACAGCTCCAGGCTTCTTCAAGTGGAGCAAAAGCCGGCCCTGTGCTAG 584
 QY 805 CCGCTGTGTGTGACTTCTTCCCTGCGCCAGAGATCAGCTTACCCCAAGACACTGCAAGCTGC 864
 Db 585 CACCTGCTGTGTGACTTCTTCCCTGCGCCAGAGATCAGCTTACCCCAAGACACTGCAAGCTGC 644
 QY 865 TCTGCGGTGAGCAGCCCTTGTGATGATGAGAGAGCGCTGCGCTGCAAGCGCTGTGCTG 924
 Db 645 TGTGCACTCAACAGCCCTTGTGAGGTGTGAGAGAGCGCTGCGCTGCTACAGCAGCTGTG 704
 QY 925 TACTCAAGAGGTGCTTCTTCTGAGCTGAGGCGCTTATCCACTACTCAGCAGACCTT 984
 Db 705 TGTCTCAAGAGGTGCGCTTCTTCAACCTGAGTCCCTTACCGGCTGTGTAAGACCCCT 764
 QY 985 CCTCAACCTGCAAGCTGTGCACTGTGCTGAGAGCCCGGCGCTGTGCTGATCCCGGG 1044
 Db 765 CCTCAACCTGCAAGCTGTGCACTGTGCTGAGAGCCCGGCGCTGTGCTGATCCCGGG 824
 QY 1045 AGCAGACCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1104
 Db 825 AACGCAAAAGGAGATCTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 884
 QY 1105 CGATCAAG 1164
 Db 885 AACTCAAG 944
 QY 1165 AACTCAAG 1224
 Db 945 AGATCTCAAG 1004
 QY 1225 GGTATGAG 1284
 Db 1005 GGTATGAG 1064
 QY 1285 GGTATGAG 1344
 Db 1065 GATGGAATTTCTTGGCCCATCTTCAAGACCTGGGTGATTAACATCACCAGAGAGAG 1124
 QY 1345 TGGGTGAGATGCTTCCACTACTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1404
 Db 1125 TGGGTGAGATGCTTCCACTACTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1184
 QY 1405 GGTGATCTTACCTTACGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1464
 Db 1185 GCTGGCTTCTTGGCTTATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1244
 QY 1465 ATTTGCTGGATTAATCTCAGATCTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1524
 Db 1245 ATTTGCTGGATTAATCTCAGATCTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1304
 QY 1525 TTCTGTCTCTCTCCACTATCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1584
 Db 1305 TTCTGTCTCTCTCTCCACTATCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1364
 QY 1585 CACCCCTTGGTCCAGCTTACAGC 1610
 Db 1365 --CCACTGTGTGACGCTCAGTAC 1388

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 10:39:13 ; Search time 478.467 Seconds

(without alignments)
10866.196 Million cell updates/sec

Title: US-09-645-078-3

Sequence: 1 gggcactcaactacactg.....ccgtacacatttttaag 1926

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_19Jun03:*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*

2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*

3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*

4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*

5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*

6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*

7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*

8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*

9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*

10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*

11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*

12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*

13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*

14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*

15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*

16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*

17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*

18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*

19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*

20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*

21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*

22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*

23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*

24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1926	100.0	1926	20	AA220793
2	751.2	39.0	1333	24	AA516947
3	741.6	38.5	1979	22	AAK94229
4	741.6	38.5	2032	20	AA220792
5	717	37.2	2065	21	AA294211
6	468.8	24.3	877	22	AAK91803
7	468.8	24.3	877	22	AAK93921
8	429.2	22.3	2968	21	AAK76156

9	370.6	19.2	1647	24	AA224670
10	370.6	19.2	1694	22	AA202700
11	370.6	19.2	2544	24	ABN89506
12	370.6	19.2	160552	22	AA202697
13	355.6	18.5	1937	22	AA516948
14	355.6	18.5	1989	22	AA202698
15	354	18.4	2044	22	AA202699
16	354	18.4	2044	22	AA202699
17	304.2	15.8	48436	24	ABN89533
18	118.6	6.2	668	24	ABQ21506
19	118.6	6.2	668	24	ABQ21507
20	114	5.9	1458	19	AAV36418
21	114	5.9	2190	24	AA41280
22	114	5.9	2415	25	ABX08787
23	110.6	5.7	7099	25	ACC46083
24	107.4	5.6	2156	19	AAV21200
25	98.4	5.1	2354	18	AAK5037
26	96.2	5.0	524	24	ABQ21510
27	96.2	5.0	524	24	ABQ21511
28	95.8	5.0	3029	22	AAH17932
29	95.4	5.0	524	24	ABQ21512
30	95.4	5.0	524	24	ABQ21513
31	94	4.9	2393	24	AA561699
32	94	4.9	2409	20	AAK87821
33	88.6	4.6	2150	20	AAK87820
34	86.8	4.5	668	24	ABQ21508
35	86.8	4.5	668	24	ABQ21509
36	82	4.3	2165	25	AB221794
37	79.8	4.1	576	24	ABQ45408
38	79.8	4.1	576	24	ABQ45409
39	65	3.4	65	24	ABN54495
40	63.2	3.3	576	24	ABQ45406
41	63.2	3.3	576	24	ABQ45407
42	52.4	2.7	731	20	AA224563
43	52.4	2.7	731	21	AAK5802
44	52.4	2.7	731	24	ABQ92207
45	52.4	2.7	731	24	ABL49021

ALIGNMENTS

RESULT 1	
AA220793	
ID	AA220793 standard; DNA; 1926 BP.
AC	AA220793;
DT	08-DEC-1999 (first entry)
XX	
XX	Mouse glycosyl sulfotransferase-3 coding sequence.
DE	
XX	
KW	Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
KW	selectin binding interaction; inflammation; lymphocyte homing; mouse;
KW	secondary lymph organ; ss.
XX	
OS	Mus sp.
XX	
XX	W09949018-A1.
XX	
XX	30-SEP-1999.
XX	
XX	26-FEB-1999; 99WO-US04316.
XX	
XX	20-MAR-1998; 98US-0045284.
XX	12-NOV-1998; 98US-0190911.
PA	(REGC) UNIV CALIFORNIA.
PA	(SYNT) SYNTAX USA INC.
XX	
PI	Blstrup A, Rosen SD, Tangemann K, Hemmerich S;
XX	
DR	WPI; 1999-580442/49.

DR P-PSDB: AAY39919.
 XX Human and murine glycosyl sulfotransferase 3 and related
 PT polynucleotides
 XX
 PS
 XX Claim 4: Fig 3; 59pp; English.

This sequence encodes the mouse glycosyl sulfotransferase-3 (GST-3) of the invention. The nucleic acid sequences, probes and primers derived from these, proteins and antibodies are useful in detecting homologues. The sequences, antibodies and methods are useful in the diagnosis and treatment of diseases associated with selectin binding interactions, including conditions associated with or resulting from the homing of leukocytes to sites of inflammation and the normal homing of lymphocytes to secondary lymph organs.

Sequence 1926 BP; 426 A; 520 C; 484 G; 496 T; 0 other:

Query Match 100.0%; Score 1926; DB 20; Length 1926;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1926; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

  QY 1 GGGCATCTACTTACCTGTGTGAGAGACAGCTTTGGCTACAAAGGCGACACTCT
  DB 1 GGGCATCTACTTACCTGTGTGAGAGACAGCTTTGGCTACAAAGGCGACACTCT 60
  QY 61 GTGAGGGGTGTGAAAGGTGTGGGTGTGGCAGAACTCCCTATAGTAAATGTCTG 120
  DB 61 GTGAGGGGTGTGAAAGGTGTGGGTGTGGCAGAACTCCCTATAGTAAATGTCTG 120
  QY 121 GTTAGATATTCGGTGGTGTGTGATGAGAGAAAGCCAGAGGGGTAAAGACT 180
  DB 121 GTTAGATATTCGGTGGTGTGTGATGAGAGAAAGCCAGAGGGGTAAAGACT 180
  QY 181 TGTACATAGATGAGAGATGATCCAAATTAAGACAGAACTTTGACAGAGGATCTGA 240
  DB 181 TGTACATAGATGAGAGATGATCCAAATTAAGACAGAACTTTGACAGAGGATCTGA 240
  QY 241 CAAATACTGTGAGAAATGTTTGTGTTTGTGTTTGTGTTTGTGTTTACATGCTC 300
  DB 241 CAAATACTGTGAGAAATGTTTGTGTTTGTGTTTGTGTTTGTGTTTACATGCTC 300
  QY 301 CTGTGATGGAGATCCAGAGAGCCGAGAGGTAGATGCTTAACAACCTACACAGCCCA 360
  DB 301 CTGTGATGGAGATCCAGAGAGCCGAGAGGTAGATGCTTAACAACCTACACAGCCCA 360
  QY 361 TCCCTCTGCTTCTCTTCAAGTCTTCTCTCTCTCCGACAGATGATGCTTGAAGA 420
  DB 361 TCCCTCTGCTTCTCTTCAAGTCTTCTCTCTCTCCGACAGATGATGCTTGAAGA 420
  QY 421 AAGGAGAGCTGCTGATGTTCCGTGGGTCCAGGTCACTGTTAGCTCTTCATCATTA 480
  DB 421 AAGGAGAGCTGCTGATGTTCCGTGGGTCCAGGTCACTGTTAGCTCTTCATCATTA 480
  QY 481 TGTCCGTCCACAGACCTTTCCACAGAGGAGAGTCCAGAGGCCCGGCAATGTCGTG 540
  DB 481 TGTCCGTCCACAGACCTTTCCACAGAGGAGAGTCCAGAGGCCCGGCAATGTCGTG 540
  QY 541 TGTCTCTCTCTGCGGGGTCCAGAGTCTTTTGTGGGACAGCTTTTCCGAGACCCGG 600
  DB 541 TGTCTCTCTCTGCGGGGTCCAGAGTCTTTTGTGGGACAGCTTTTCCGAGACCCGG 600
  QY 601 ATGTGTCTACTGATGAGAGCTGCTGAGAGTGTGATGACTTTACACACAGACAG 660
  DB 601 ATGTGTCTACTGATGAGAGCTGCTGAGAGTGTGATGACTTTACACACAGACAG 660
  QY 661 CTGGAAGTGCACATGGCTGTGGGGATCTTCTGGTCCGTCTCCGTGTGACATTA 720
  DB 661 CTGGAAGTGCACATGGCTGTGGGGATCTTCTGGTCCGTCTCCGTGTGACATTA 720
  QY 721 GCGCTTTGATGCCATACATGAACCCAGAGCCCGGAAACAGTCCAGCCTTCCAGTGG 780
  DB 721 GCGCTTTGATGCCATACATGAACCCAGAGCCCGGAAACAGTCCAGCCTTCCAGTGG 780

```

```

  QY 781 AGCAAAAGCCGGCCCTGTGCTCAGCGCCTGTGTGTACTTCTCCTGCCACGATCA 840
  DB 781 AGCAAAAGCCGGCCCTGTGCTCAGCGCCTGTGTGTACTTCTCCTGCCACGATCA 840
  QY 841 GCTACCCCAAGCATCAGCTGCTGGGTCAGAGCCCTTGTATGTGTGAGAAAG 900
  DB 841 GCTACCCCAAGCATCAGCTGCTGGGTCAGAGCCCTTGTATGTGTGAGAAAG 900
  QY 901 CCTGCGGCTCAGCGCTTCGTGATCTCAAGAGGTGGTCTTCTACGCTCAGAGCC 960
  DB 901 CCTGCGGCTCAGCGCTTCGTGATCTCAAGAGGTGGTCTTCTACGCTCAGAGCC 960
  QY 961 TCTATCTACTACTACGAGACCTTCCCTCAACCTGACGCTGTGACCTGTCGAGAC 1020
  DB 961 TCTATCTACTACTACGAGACCTTCCCTCAACCTGACGCTGTGACCTGTCGAGAC 1020
  QY 1021 CCGGGCCGTGTCCCATCCCGGAGACACACCATTAACATCATGTGTGACATCA 1080
  DB 1021 CCGGGCCGTGTCCCATCCCGGAGACACACCATTAACATCATGTGTGACATCA 1080
  QY 1081 TTGTGCTAGGAGCATTTTGGAAACGATCAAGAGAGAACCCATATATGACATGA 1140
  DB 1081 TTGTGCTAGGAGCATTTTGGAAACGATCAAGAGAGAACCCATATATGACATGA 1140
  QY 1141 AGATCATCTGCAAAAGCCAGGTGACATGATCAAGGACATCCAAACCTCCTGAAGCTC 1200
  DB 1141 AGATCATCTGCAAAAGCCAGGTGACATGATCAAGGACATCCAAACCTCCTGAAGCTC 1200
  QY 1201 TGCAGAGGCTTACCTGTCCTGAGATGAGAGACTGTTGGGGACCCCGGCCACA 1260
  DB 1201 TGCAGAGGCTTACCTGTCCTGAGATGAGAGACTGTTGGGGACCCCGGCCACA 1260
  QY 1261 CGACGACATATATATTTGTGGGGTGTGATTTTGGCCGACCTCCAAACATGGGTT 1320
  DB 1261 CGACGACATATATATTTGTGGGGTGTGATTTTGGCCGACCTCCAAACATGGGTT 1320
  QY 1321 ACAATGTACCCGCGGCAAGGGCATGTGATGATGCTCTTCCATTAACGCCAGAACG 1380
  DB 1321 ACAATGTACCCGCGGCAAGGGCATGTGATGATGCTCTTCCATTAACGCCAGAACG 1380
  QY 1381 CCCTCAAGCTCTCAGGGGTGGGCTGTGGCTTACCTTACGAAAGGTTTCCAGACTTC 1440
  DB 1381 CCCTCAAGCTCTCAGGGGTGGGCTGTGGCTTACCTTACGAAAGGTTTCCAGACTTC 1440
  QY 1441 AAGATGCTGCGGTGAGGCTATGATGATTTCTGGGATACCTCCAGGTCAAGTCAACAAG 1500
  DB 1441 AAGATGCTGCGGTGAGGCTATGATGATTTCTGGGATACCTCCAGGTCAAGTCAACAAG 1500
  QY 1501 AACCAAGCAACCTGTCCCTGATCTTCTCTCCCATATCTTGGGGCAGGTCTTC 1560
  DB 1501 AACCAAGCAACCTGTCCCTGATCTTCTCTCCCATATCTTGGGGCAGGTCTTC 1560
  QY 1561 GAGAAGGTTAAGAGGTGTGTCTGACACCCCTTGGTCCAGCCTTACTACATTAACGC 1620
  DB 1561 GAGAAGGTTAAGAGGTGTGTCTGACACCCCTTGGTCCAGCCTTACTACATTAACGC 1620
  QY 1621 ACAGAAAGCTTAAGATATTAACCAAACTGATGAGTCCCTTCTCCAGCCCAAGCAGAG 1680
  DB 1621 ACAGAAAGCTTAAGATATTAACCAAACTGATGAGTCCCTTCTCCAGCCCAAGCAGAG 1680
  QY 1681 GGTCTTGTGTATATCTATGATGATGCTTACCTTACAGTACGACCTTAAAGCAAGCAAT 1740
  DB 1681 GGTCTTGTGTATATCTATGATGATGCTTACCTTACAGTACGACCTTAAAGCAAGCAAT 1740
  QY 1741 ATCTTCTGCTTGAATAATACTAGGAACCTTAAGAGGCCCTTGAACCTGTCAAGCAAG 1800
  DB 1741 ATCTTCTGCTTGAATAATACTAGGAACCTTAAGAGGCCCTTGAACCTGTCAAGCAAG 1800
  QY 1801 ACTTCTGTAACTTGGCCTTCTTACCTGTGATACCTTGAAGACTCGGTGTGAGAGCA 1860
  DB 1801 ACTTCTGTAACTTGGCCTTCTTACCTGTGATACCTTGAAGACTCGGTGTGAGAGCA 1860

```


QY	1297	TGCCCACTCCAAACATGGGTTTACATATGACCCGGGCAAGGGATGGGTACGATG	1356
Db	988	TGCCCATCTTACACCTGGGTGCATATCATCACCCAGGCAAGGGATGGGTATCCACG	1047
QY	1357	CTTTCATCTACTAACGCCAGGAAAGCCCTCAACGTCTCTCAGGCGTGGCGTTGGTCCTTAC	1416
Db	1048	CTTTCACACACAAATGCGCAGGGATCCCTTATATGTCTCCAGGCTTGGCGGTCTTTTGC	1107
QY	1417	CTTACGAAAAGTGTCCCGAGCTTCAATATGCGTCCGGTGAAGCATATGATTTGGTGGGAT	1476
Db	1108	CCATATGAAAAGGTCTTCGACTTCAGAAAGCTTGGCGATGCCATTAATTTGGTGGGCT	1167
QY	1477	ACCTCCAGGTCCAGTCTCCACACAGAAACAAAGCAACTGTCCCTGGATCTTGTCTCTCT	1536
Db	1168	ACGCCACAGTCCAGTCTGAAACAAAGAAAGAAACCTGTGCTGGATCTTGTCTCTACT	1227
QY	1537	CCCATATCTTGGGGCAGGCTTTCGAGAAAGGTCTGAAGAGCTGTGTGCACCCCTTGGTT	1596
Db	1228	GGACTGTCCCTGAGCAATCCATTAAGAGGGTTGAGAAGGCTTGTCTG -- CCACCTGGTG	1285
QY	1597	CCAGCCTTAGTCAC	1610
Db	1286	TCAAGCTCAGTCAC	1299

RESULT 3	
ID	AAK94229 standard; cDNA; 1979 BP.
AAK94229	
AAK94229;	
06-NOV-2001	(first entry)
Human full-length cDNA, SEQ ID NO: 2816.	
Human: full length cDNA; cDNA synthesis; oligo-capping; ss.	
Homo sapiens.	
EP130094-A2.	
05-SEP-2001.	
07-JUL-2000; 2000EP-0114089.	
08-JUL-1999; 99JP-0194486.	
11-JAN-2000; 2000JP-0118774.	
02-MAY-2000; 2000JP-0183765.	
(HELT-) HELIX RES INST.	
Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;	
Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;	
WPI: 2001-524255/58.	
P-PSDB; AAK93309.	
830 Primers useful for synthesizing full length cDNA clones and their	
use in genetic manipulation -	
Claim 8; SEQ ID NO 2816; 1380bp + sequence listing; English.	
The invention relates to primers for synthesizing full length cDNA	
clones. 830 cDNA molecules encoding a human protein have been	
isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA	
molecules have been determined. Primers for synthesizing the full length	
cDNA are useful for clarifying the function of the protein encoded by	
the cDNA. The full length clones were obtained by construction of full	
length enriched cDNA libraries that were synthesised by the oligo-capping	
method. The primers enable the production of the full length cDNA easily	
without any special methods. The present sequence is a full length	
human cDNA of the invention.	

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
S0 Sequence 1979 BP; 451 A; 542 C; 474 G; 512 T; 0 other;

Query Match	38.58;	Score 741.6;	DB 22;	Length 1979;
Best Local Similarity	76.48;	Pred. No. 4.8e-203;		
Matches 937; Conservative	0;	Mismatches 284;	Indels 5;	Gaps 2

QY	388	TCCTCCCTTCGCGAGATGATGCGTTGGAAGAAAGGAGAGCGTGTATGTTCCGGGT	447
Db	52	TCCTCCACTTACGACCAATGCTACTGCTTAAAAAATAGACTCTGCTGTTTCGTGTT	111
QY	448	CCGAGGTCATGTTGTAGCTCTCTTCATTCATATGTCGTCACAGACCTTTCACG-	506
Db	112	CCGAGATGGCATCTTGCTCTATTCCTTCACATGTACAGCCACAAATCAGCTCCGCT	171
QY	507	--AGGAGGAACTCCAGAGGCCCGGTGCATGTGCTGTCTGTTCTTCGTGGGGTACAGAT	564
Db	172	CTATGAAGGACAGGCCCGGAGCAAGTCAGTGGTTCCTGCTTCCTGGGGCTTGCT	231
QY	565	CCCTCTTTTGTGGAGACAGCTTTTGGGGACAGACCCGATGTGTCTTACCTGATGGAGCGT	624
Db	232	CTTCTTTTGTGGGAGAGCTTTTGTGGGACAGACCCAGATGTTTCTACTGTATGAGCCCG	291
QY	625	CCTGGCATGTGTGATGACTTTCACACAGACAGACCTGTGAAGCTGCATAGCTGTGC	684
Db	292	CCGTGGACAGTGTGATGTGACTTCAAGACAGAGACGCGCTGGATGTGCACATGGCTGTGC	351
QY	685	GGGATCTTCTGCGTTCCGTCCTTTCCTGTGTGATATGATAGAGCTCTTTATGCTCATGAAC	744
Db	352	GGGATCTGTGATACGGGCGCTCTCTCTGTGTGCGACATGAGGCTCTTGTATCTCAATGAAC	411
QY	745	CAGAGCCCCGGAAACAGTCCAGACCTCTTCCAGTGGGACAAAGCCGGGCGCTGTCTAG	804
Db	412	CTGGTCCCCGGAGACAGTCCAGACCTCTTTCAGTGGGAAACAGCCGGGCGCTGTCTAG	471
QY	805	CGCCTGTGTGACTTCTTCCCTGCGCCACGAGATCAGCTCACCCAAAGCACTGCACAGCTGC	864
Db	472	CACCTGCCTGTGACATCATCTCCACAAAGATGAATATCATCCCGGGCGTCAACAGGCTCC	531
QY	865	TCCTCGGTCACAGACCTTTTATATATGATGGGAAAGAGCCTGCCGCTCTACAGGCTTCGTGG	924
Db	532	TGTCCAGTCAACAGCCCTTTTGAAGGTGGTGGAAAGAGCCCTGCGCTCCATACAGCCAGCTGG	591
QY	925	TACTCAAGGAGAGTGGGTTTCTTCAGACCTGCAGAGCCCTTATCCACTACTACAGACCTT	984
Db	592	TGCTCAGAGAGTGGCTTCTTTCACACCTGCAGTCCCTTACCGCTGCTGAAAGACCCCT	651
QY	985	CCCTCAACCTGCAGCTGTGCACCTGTGTCCAGACCCCGGGCGCTGTTCGCATCCCGGG	1044
Db	652	CCCTCAACCTGCATATGCTGTGCACCTGTGTCCGGAGCCCGGGCGCTGTTCGGTCCCGAG	711
QY	1045	AGCACACCACTATGAACTCATGTGTGACAGTCAATATGTGCTAGGGACAGCATTTTGGAA	1104
Db	712	AACCCACAAAGGAGATCTCATGTATGTCAAATCGCATTTGATATGGGACAGCATGACAA	771
QY	1105	CGATCAAGGAGGAAGACCGCCATTTATGCAATGATCATGTGCAAAAAGCCAGGAG	1164
Db	772	AACCTCAAGGAAGAGACCAACCCCTCATATGATGATGATCAAGTCAATCTGCGCAAAAGCCAGCTGG	831
QY	1165	ACATAGTCACAGGCCATCCAAACCCCTCCCTGAAAGCTCTCAGACAGCGCTACCTGTTCTCTGA	1224
Db	832	AGATCTACAGAACCATCCAGTCTTGTGCCCAAGGCCCTCAGAAAGCTTACCTGCTTGTGCG	891
QY	1225	GGATGAGAGACTGTGCGGGACCCCTGGGCCACAGACGACAGACTATATAATTTGTGG	1284
Db	892	GCTATGAGAGACTGGCTGAGACCCCTGCGGCCACAGACTTCCGAAATGATATAATTCGTGG	951
QY	1285	GGTTGGATTTTTCGCCCACTCCAAACATGGGTTTACATATGTACACCCGGCGCAAGGGCA	1344
Db	952	GATTGGAATCTCCCATCTTACAGACTGGGTGCATACATCAACCCGAGGCAAGGGCA	1011

QY 1465 ATTGCTGGGATACCTCCAGGTCAGATCTCAACAGACAAAGCAACCTGCTCCGATC 1524
 Db 1245 ATTTGCTGGGATACCTCCAGGTCAGATCTCAACAGACAAAGCAACCTGCTCCGATC 1304
 QY 1525 TTCTGCTCCGATACCTGAGGTCAGATCTCCAGAGAGGTTAAGGAGGTCGTCG 1584
 Db 1305 TTCTGCTCCGATACCTGAGGTCAGATCTCCAGAGAGGTTAAGGAGGTCGTCG 1364
 QY 1585. CACCCCTGGTCCAGACCTTACGAC 1610
 Db 1365 --CCACCTGGTCCAGACCTTACGAC 1388

RESULT 5
 AA294211
 ID AA294211 standard; cDNA; 2065 BP.
 XX AA294211;
 DT 19-JUN-2000 (first entry)
 XX
 XX Human transferase TRNSFS-11 cDNA clone 2617407C1.
 XX
 XX Transferase: TRNSFS-11; human; antitumor; cell proliferation;
 KM inflammation; gastrointestinal disorder; developmental disorder;
 KM genetic disorder; neurological disorder; reproductive disorder;
 KM smooth muscle disorder; immunological disorder; gene therapy;
 KM diagnosis; N-acetylglucosamine 6-O-sulfotransferase; s9.
 XX
 XX Homo sapiens.
 OS
 FH
 FT Key Location/Qualifiers
 FT CDS 174..1334
 FT /tag= a
 PN WO200014251-A2.
 XX
 PD 16-MAR-2000.
 XX
 PF 09-SEP-1999; 99MO-US20989.
 XX
 PR 10-SEP-1998; 98US-0150657.
 PR 04-NOV-1998; 98US-0186779.
 PR 11-MAY-1999; 99US-0133642.
 XX
 XX (INCY-) INCYTE PHARM INC.
 PA
 PI Tang YF, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;
 PI Hillman JL, Azimzal Y;
 DR MPI, 2000-256996/22.
 DR P-PDB; AAT79219.
 XX
 XX Human transferase proteins useful for preventing, diagnosing and
 PT treating cancers and developmental, gastrointestinal, genetic,
 PT immunological, neurological, reproductive and smooth muscle disorders -
 XX
 PS Claim 9; Page 104-105; 113pp; English.
 XX
 XX The present sequence is that of cDNA clone 2617407C1 encoding
 CC human transferase TRNSFS-11 (see AAY79219), 1 of 15 claimed human
 CC transferase proteins of the invention (see AAY79209-23). The clone
 CC was isolated from gall bladder cDNA library GBLN0701. TRNSFS-11
 CC is expressed in dermal and gastrointestinal tissues,
 CC especially those associated with inflammation and cell
 CC proliferation. It shows homology to mouse N-acetylglucosamine
 CC 6-O-sulfotransferase. A polynucleotide comprising nucleotides
 CC 264-323 or 1272-1331 of the present sequence can be used as a DNA
 CC probe. The new human transferases and polynucleotides can be used
 CC in the diagnosis, prevention and treatment (including gene therapy
 CC and antisense therapy) of cancer, developmental disorders,
 CC gastrointestinal disorders, genetic disorders, immunological
 CC disorders, neurological disorders, reproductive disorders, and

CC smooth muscle disorders.
 XX Sequence 2065 BP; 466 A; 575 C; 491 G; 533 T; 0 other;
 SQ

Query Match 37.2%; Score 717; DB 21; Length 2065;
 Best Local Similarity 76.2%; Pred. No. 6.2e-196;
 Matches 935; Conservative 0; Mismatches 285; Indels 7; Gaps 4;

QY 388 TCTCTCTTCCGACGATGATGCTTGTAGAAAGGAGGCTGATGTTCTCGGTT 447
 Db 157 TCTTCACCTTCAGACAAATGCTACTGCTCAAAAAAATGAAGCTCGTGTCTGTTT 216
 QY 448 CCCAGGTATCGTTGAGCTCTCTCATATATGTCCTCCACAGACACTTCCAG- 506
 Db 217 CCCAGATGACCATCTTGCTCATATGTCCTCATATGTCCTCCACAGACACTTCCAGT 276
 QY 507 --AGGAGAGAGTCACAGAGAGCCCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 564
 Db 277 CTATGAAAGCACAGCCCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 336
 QY 565 CCTCTTTTGGGACAGCTTTTGGGACAGCCGAGATGTTCTACTGATGAGCCCTG 624
 Db 337 CTCTTTTGGGAGGAGCTTTTGGGACAGCCAGATGTTCTACTGATGAGCCCG 396
 QY 625 CCTGGCATGTCGATGACTTTACACAGACACACCTGGAAGCTGCACATGGCTGTC 684
 Db 397 CCTGGCATGTCGATGACTTTACACAGACACACCTGGAAGCTGCACATGGCTGTC 456
 QY 685 GGGATCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 744
 Db 457 GGGATCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 516
 QY 745 CAGGCCCCGGAACAGTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 804
 Db 517 CTGCTCCCGGAGACAGTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 576
 QY 805 CGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 864
 Db 577 CACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 635
 QY 865 TCTGCGTACAGACGCTTTTG-ATATGTCGAGAAAGGCTGCTGCTGCTGCTGCTGCTG 923
 Db 636 TGTGAGTCAACAGCCCTTTGAAGTTGTGGAAGAAAGCTGCTGCTGCTGCTGCTGCTG 695
 QY 924 GTACTCAGAGAGTGGCTTTTCTACGCTGACGAGCCCTTATCACTACTACGAGCCT 983
 Db 696 GTGCTCAAGAGTGGCTTTTCTCAACAGTCAAGTCCCTTACCCCTGCTGTAAGACCCC 755
 QY 984 TCCCTAACCTGACCTGTCGACCTGTCGACCTGTCGACAGAGAGCCCGGCTGTCGATCCCG 1043
 Db 756 TCCCTAACCTGATATGTCGACCTGTCGACCTGTCGACAGAGAGCCCGGCTGTCGATCCCG 815
 QY 1044 GAGCACACCACTAGACTATGATGTCAGCTATGTCATGTCGTCAGGAGCATTTGGAA 1103
 Db 816 GAACGCAAGAGAGATGTCATGATGTCAGCTATGTCATGTCGTCAGGAGCATTTGGAA 875
 QY 1104 ACGATCAAG 1163
 Db 876 AACTCAAG 935
 QY 1164 GACATGTCAG 1223
 Db 936 GAGATTCAG 995
 QY 1224 AGTATGAG 1283
 Db 996 CGCTATGAG 1055
 QY 1284 GGGTTGAGATTTTGGCCACCTCCCAACATGAGGTTTACATGTCACCCGCGAGAGGCG 1343
 Db 1056 GGGTTGAGATTTTGGCCACCTCCCAACATGAGGTTTACATGTCACCCGCGAGAGGCG 1115
 QY 1344 ATGGGTACAGAGCTTCCATATCTAAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 1403

Db 1116 ATGGTGACACAGCTTTCCACACAAATCCAGGATGCCCTTATGCTCCAGGCTGG 1175
 QY 1404 CGTTGTCCTTACCTTACGAAAGGTTTCCAGCTTCAAGTGGCTCGGGAGGCTATG 1463
 Db 1176 CGGTGCTTGGCTTGAAGGTTTCTGACTTGAAGAGCTTGGGATGCGATG 1235
 QY 1464 GATTTCGGGATACCTCCAGGTGATCTCAACAGAACAGCAAGCTTCCCTGGAT 1523
 Db 1236 AATTGCTGGGCTACCCGACGTCAGATCTGACAGAACAGCAAGCTTCTGGAT 1295
 QY 1524 CTTCCTGCTCCCTCCATCTTGGGAGGCTTCCGAGAGGTTAAGAGTCTGCT 1583
 Db 1296 CTTCCTGCTACCTGAGCTGCTCCCTGAGCAAAATCCAGTAAAGAGGCTTGGCT 1355
 QY 1584 GCACCCCTGGTCCAGCTTATGTCAC 1610
 Db 1356 G--CCACCTGGTGTACGCTCAGTCAC 1380

RESULT 6

AAK91803 standard; cDNA: 877 BP.

AAK91803;

06-NOV-2001 (first entry)

Human cDNA 5'-end sequence, SEQ ID NO: 263.

Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

Homo sapiens.

EP1130094-A2.

05-SEP-2001.

07-JUL-2000; 2000EP-0114089.

08-JUL-1999; 99JP-0194486.

11-JAN-2000; 2000JP-0118774.

PR 02-MAY-2000; 2000JP-0183765.

PA (HELI-) HELIX RES INST.

Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

WPI; 2001-524255/58.

830 Primers useful for synthesizing full length cDNA clones and their use in genetic manipulation -

Claim 2; SEQ ID NO 263; 1380bp + sequence listing; English.

The invention relates to primers for synthesizing full length cDNA clones. 830 cDNA molecules encoding a human protein have been

isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA

molecules have been determined. Primers for synthesizing the full length

cDNA are useful for clarifying the function of the protein encoded by

the cDNA. The full length clones were obtained by construction of full

length enriched cDNA libraries that were synthesised by the oligo-capping

method. The primers enable the production of the full length cDNA easily

without any special methods. The present sequence is the nucleotide

sequence of the 5'-end of a cDNA provided in the invention.

Note: The sequence data for this patent did not form part of the printed

Matches 585; Conservative 0; Mismatches 174; Indels 3; Gaps 1;

QY 388 TCCCTTCTCCAGAGATGATGCTGTGAAGAAAGGAGGCTGATGCTCGGGTT 447
 Db 52 TCTTCCACTTCCAGCAATGCTACTGCTCTAAAAAATGAATACCTCTGTTTGTTT 111
 QY 448 CCCAGTCATGCTGTGATCTCTTCAATCCATATGTCGCTCCAGACCTTCCAG- 506
 Db 112 CCCAGATGCGCATTTGGCTCTATTTCCATGATGACAGCAACATGATGCTCCTGT 171
 QY 507 -AGGAGAGATCCAGAGGCGCGCATGCTGCTGCTCTCTTCTGCGGTGAGAT 564
 Db 172 CTATGAAGGACACCGGAGGCGCATGACGCTGTGTTCTCTGCGGTGAGAT 231
 QY 565 CTTCTTTTGGGAGACCTTTTGGGAGACCCCGGATGTTTCTTACTGATGAGCTG 624
 Db 232 CTTCTTTTGGGAGACCTTTTGGGAGACCCCGGATGTTTCTTACTGATGAGCTG 291
 QY 625 CTTGCAATGCTGATGATCTTTCACAGCAGCAGCCTGGAAGCTGACATGCTGTC 684
 Db 292 CTTGCAATGCTGATGATCTTTCACAGCAGCAGCCTGGAAGCTGACATGCTGTC 351
 QY 685 GGGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 744
 Db 352 GGGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 411
 QY 745 CAGGCCCCCGGAACACTTCCAGCTTCCAGTGGGAGCAAGCGGCGCTGCTGCTG 804
 Db 412 CTTGCTCCCGGAGACAGTCCAGCTCTTTCAGTGGGAGCAAGCGGCGCTGCTGCTG 471
 QY 805 CGCCTGTGTGATCTTCCCTGCTCCAGCAGATCAGCTCACCAGCACTGCAAGCTG 864
 Db 472 CACCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 531
 QY 865 TCTGCGGTGACAGCCTTTGATATGCTGAGGAAGCGCTGCTGAGGCTTCTGCTG 924
 Db 532 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 591
 QY 925 TACTCAAGAGGAGGCTGTTTCTGAGCCTGAGGCGCTGATCCAGTACTGAGACCTT 984
 Db 592 TGTCTCAAGAGGAGGCTGTTTCTGAGCCTGAGGCGCTGATCCAGTACTGAGACCTT 651
 QY 985 CCTTCAACCTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1044
 Db 652 CCTTCAACCTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 711
 QY 1045 AGCAGACCCAGCATGAACTCATGCTGACAGTCAATTTGCTGAGGAGCATTTGAAA 1104
 Db 712 AACGCAAGAAAGGAGATCTCATGATTGCAATCCATTTGATGAGGAGCAATGAGCAA 771
 QY 1105 CGATCAAG 1146
 Db 772 AACTCAAGAAAG 813

RESULT 7

AAK93921 standard; cDNA: 877 BP.

AAK93921;

06-NOV-2001 (first entry)

Human cDNA clone representative sequence, SEQ ID NO: 2381.

Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

Homo sapiens.

EP1130094-A2.

05-SEP-2001.

Query Match. 24.3%; Score 468.8; DB 22; Length 877;

Best Local Similarity 76.8%; Pred. No. 1.6e-124;

Query Match	Best Local Similarity	Matches	Score	DB	Length
24.3%	76.8%	0	468.8	22	877
Conservative	Pred. No. 1.6e-124	Mismatches 174			
		Indels 3			
		Gaps 1			
Sequence 877 BP, 183 A, 266 C, 221 G, 203 T, 4 other;					
Example 11: SEQ ID NO 2381, 1380bp + sequence listing; English.					
The invention relates to primers for synthesizing full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5' and 3' ends of the cDNA molecules have been determined. Primers for synthesizing the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence was used as the representative sequence from a human clone which was used in homology searches to identify the clone.					
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO.					
Query Match	Best Local Similarity	Matches <td>Score</td> <td>DB</td> <td>Length</td>	Score	DB	Length
24.3%	76.8%	0	468.8	22	877
Conservative	Pred. No. 1.6e-124	Mismatches 174			
		Indels 3			
		Gaps 1			
Sequence 877 BP, 183 A, 266 C, 221 G, 203 T, 4 other;					
Example 11: SEQ ID NO 2381, 1380bp + sequence listing; English.					
The invention relates to primers for synthesizing full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5' and 3' ends of the cDNA molecules have been determined. Primers for synthesizing the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence was used as the representative sequence from a human clone which was used in homology searches to identify the clone.					
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO.					

QY	925	TACCAAGAGAGTGGCTTTCTCAGAGCCGACAGCCCTATATCCACTACTCAGAGACCTT	984
Db	592	TGCTCAGAGAGTGGCTGCTTCTTCAACCTCGACAGTCCCTTACCCTGCTGAAAGACCTT	651
QY	985	CCCTCAACCTTCACGCTGCTGCACCTGTGTCCGAGACCCCGGGCGCTGTTCCGATCCCGGG	1044
Db	652	CCCTCAACCTTCATATGCTGCACCTGCTCCGGAGACCCCGGGCGCTGTTCCGATCCCGAG	711
QY	1045	AGCAGACACCATATGAACTCATATGTTGTGACATCATATGTGCTGAGGAGACATTTGGAAA	1100
Db	712	AAGCCACAGAGAGGAGATCTCATGTAGTGAACATCCGATTGNGATGGGGCAACATGAGCAAA	771
QY	1105	CGATCAAGAGAGAGAGACAGCCCTATTTATGTCATGAATGAATGA 1146	
Db	772	AACTCAGAGAGAGAGACCAACCCCTACTATGTGTGATGACAGGCA 813	
RESULT 8			
ID	AAC76156	standard; cDNA; 2988 BP.	
AC	AAC76156;		
DT	08-FEB-2001	(first entry)	
DE	Human ORFX ORF1711 polynucleotide sequence SEQ ID NO:3421.		
XX	Human; open reading frame; ORFX; detection; cyostatic; hepatotropic;		
KW	vulnerable; antiparkinsonian; antiparkinsonian; nootropic; neuroprotective;		
KW	anticonvulsant; osteopathic; antidiabetic; immunosuppressant; cardiant;		
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;		
KW	hypotensive; dematological; immunosuppressive; antinflammatory;		
KW	antiviral; antibacterial; antifungal; antihemetic; antihypertoid;		
KW	antianaemic; gene therapy; cancer; proliferative disorder; hypertension;		
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;		
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;		
KW	cholesterol ester storage; systemic lupus erythematosus; infection;		
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;		
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;		
KW	bone damage; cartilage damage; antinflammatory disease; coagulation;		
KW	thrombosis; contraceptive; ss.		
OS	Homo sapiens.		
PN	WO200058473-A2.		
PD	05-OCT-2000.		
PF	31-MAR-2000; 2000WO-US08621.		
PR	31-MAR-1999; 99US-0127607.		
PR	02-APR-1999; 99US-0127636.		
PR	05-APR-1999; 99US-0127728.		
PR	30-MAR-2000; 2000US-0540763.		
XX	(CURA-) CURAGEN CORP.		
XX	Shimkets RA, Leach M;		
XX	WPI; 2000-602362/57.		
XX	P-PsDB; AAB41947.		
XX	Novel nucleic acids and peptides derived from open reading frame X,		
XX	useful for treating e.g. cancers, proliferative disorders,		
XX	neurodegenerative disorders and cardiovascular disease -		
XX	Claim 5; Page 2597-2599; 5507Pp; English.		
XX	AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,		
XX	which represent the human ORFX open reading frames 1 to 3161. The ORFX		
XX	sequences have activities such as: cyostatic; hepatotropic; vulnerable;		
XX	antiparkinsonian; antiparkinsonian; nootropic; neuroprotective;		

The invention relates to human drug metabolising enzymes referred as DME and nucleic acid molecules encoding such enzymes. Polynucleotides of the invention are useful for assessing toxicity of test compounds and in gene therapy. Sequences of the invention are useful in the diagnosis, prevention and treatment of autoimmune/inflammatory disorders such as acquired immune deficiency syndrome (AIDS), adult respiratory distress syndrome, allergies, anaemia, atherosclerosis, asthma, autoimmune haemolytic anaemia, contact dermatitis, Crohn's disease, glomerulonephritis, Goodpasture's syndrome, gout, Grave's disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple sclerosis, osteoarthritis, osteoporosis, psoriasis, systemic lupus erythematosus, rheumatoid arthritis, scleroderma, ulcerative colitis, warts, viral, bacterial, fungal, parasitic, protozoal, helminthic infections, cell proliferative disorders such as actinic keratosis, arteriosclerosis, atherosclerosis, Duchenne and Becker dystrophy, hepatitis, Cushing's syndrome, cancers, myelodysplastic syndrome, epilepsy, endocrine disorders such as disorders of the hypothalamus and pituitary resulting from lesions such as primary brain tumours, adenomas, infarction associated with pregnancy, aneurysms, vascular malformations; eye disorders such as conjunctivitis, iritis, retinitis, glaucoma, pigmentosa; metabolic disorders such as Addison's disease, cystic fibrosis, diabetes, goitre, glycogen storage diseases, hypercholesterolaemia, hyperthyroidism, hypoglycaemia, lipid myopathies Menkes syndrome, mannosidosis, obesity, gastrointestinal disorders such as dysphagia, gastric carcinoma, anorexia, nausea, gastroenteritis, hyperebilirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental disorders. The present sequence is human DME-5 protein cDNA.

	Query Match	19.2%;	Score 370.6;	DB 24;	Length 1647;	
	Best Local Similarity	63.2%;	Pred. No.5.2e-96;			
	Matches	641;	Conservative	0;	Mismatches 359;	Indels 15; Gaps 4
OY	524	GCCCGTCATGTGCTGTGCTGTCTTCCGTGGCGGTACAGATCCCTTTTGTCGACAGCT	583			
Db	117	GC GGTCATGTGCTGTGCTGTCTCCTCGTCGGCGCTCGGGCTGTCTTGTGGCAACT	176			
OY	584	TTTTGGGACAGACC GGATGTGTTCTACTCATGATGGAAGGCTGCTGTGGCATGTGGATGAC	643			
Db	177	CTTTCACCAAGCACCCCGAGCTCTTCTACTAATGAGGCCGCGGTGGACAGGTGGAACAC	236			
OY	644	TTTCACCAAGCACACAGCCTGGAACTGTCACATGGCTGTGGGGATCTTCTCGTTCGCT	703			
Db	237	CCTGTCCAGGGGACGCGCGGACACGTCACATGGCTGTGGCGACCTGTGTCCGCTCGT	296			
OY	704	CTTCTCTGTGACATGAGCCCTTTGATGCTTACTATGAACCAAGCCCCCGGAACAAGTC	763			
Db	297	CTTCTCTGTGGACATGAGCGTTTGAAGCCTATTCTG---CCTTGGCCCGGCAACCTGTGC	353			
OY	764	CAGCCTCTTCCAGGGGAGCAAACCCGGGGACCTGTGCTCACAGCGCCTGTGTGACATTCCT	823			
Db	354	CGACCTCTTCCAGTGGGGCCGTGAGCCCGTGCACCTGTGCTGGCACCCCGCGCTGAGTGCCTT	413			
OY	824	CCCTGGCCACAGATTCAGCTTCACCCAAGCACTGGAAGCTGTGCTCGGTACAGACCCCTT	883			
Db	414	TCCCAGAGGGCCATCAAGACGAGAGGGCGTGTGCAAGCACTGTGTCCGGCGGCAATCCTT	473			
OY	884	TGATATGTGTGAGAGAGCGCTGCGCTCTCACAGGCTTCTGTGTACTCAAGAAGTGCCTTT	943			

Dd		.474	CACCCCTGCGGGAGGCCCTGCCTCCTACAGCCACTGGTGGCTCAAGAGAGTGTCGCTT	533		
Oy		944	TCTCAGCCTGAGAGCCCTCTCATCCACTACTCACAGACCCTTTCCCTCAACCTGCACGTGT	1003		
Dd		534	CTTCAACCTGGAGGGTCTCTACCCGCTCTCAGGAGACCCTCCGCTCAACCTACGATAGT	593		
Oy		1004	GCACCTGGTCGAGACCCCCGGCCGCTGTTCCGATTCCCAGGAGCACACCACTATGAACCT	1065		
Dd		594	GCACCTGGTGGCGCACCCGGCGGCCCTCTCTGCTCCCGGAGACGACAGCCAAAGCTCT	653		
Oy		1064	CATGGTTGACAGCATATATTGTCTAAGGACACATTTGGAAACGATCAAGAGAAGACCA	1123		
Dd		654	GCGCGGTGACAACGGCATCTGTGTGGACCAA---GCGCAGTGGGGGAGGCCGACCC	710		
Oy		1124	GCCCTATTATGCATGAAGATCATCTGCAAAAAGCCAGTGGACATATGTCAAAGCCCATCA	1183		
Dd		711	CGGCTGCGCGTGGTGGCGGAGGTGGCCGTAGCCACGTAACGATTCGCGGAGGCGCCAC	770		
Oy		1184	AACCTCCCTCCATAAGC---TCTGACAGACAGCCGTACCTGTTCCGTGAGGTATGAGACCTGT	1240		
Dd		771	ACTCAAGCGCGCACCCCTTCTGCGCGGCGCGCTACCGCCTGGTGGCTTCGAGAACGTGC	830		
Oy		1241	TGCGGACACCCCTGGCGCCAGACAGCACACTATATTAATTGTGGGGTTGGATTTTTTCC	1300		
Dd		831	GCGGAGAGCGCTGGCAGAAATCCGTGGCTCTACGCTTCACTGGGCTCACTCTCACGCC	890		
Oy		1301	CCACCTCCCAACATAGGGTTTACATATATGCACCCGGCGGCAAGGAGCATAGGTCAGC-----A	1354		
Dd		891	ACAGCTGAGGCGCTGTGATCCATATACATACCCACGAGATCTGGACCTGGTGGCGCCGCA	950		
Oy		1355	TGCGTTCCATCTAACGCCAGAGAACGCCCTCAACGTCTCTCAGGCGTGGCGTTGGTCTT	1414		
Dd		951	AGCCTTCAGACTCTCTCCAGAGATGGGCTCAACGTCTCCAGCGCTGGCGCCCATGCGCT	1010		
Oy		1415	ACCTTAGCAAAAGGTTTCCAGCTTCAAGATGCGTGGGGTGAAGGCTATATGATTCGCGGG	1474		
Dd		1011	GCCCTTTGCCAAGATCCGCGCGCTGCAGAGACTGTGGCGTGTGCGCTGCAGCTGTGG	1070		
Oy		1475	ATACCTCCAGGCTCAGATCTCAACAAGAACAGCAAGCAACCTGTCCCTGGATCTTTCG	1529		
Dd		1071	CTACCGGCGTGTACTCTGAGGACGAGACGCAACCTGCGCTTGATCTGGTG	1125		
<hr/>						
RESULT 10						
AAD02700 standard; cDNA: 1694 BP.						
ID	AAD02700					
XX	AAD02700;					
DT	02-MAY-2001	(first entry)				
DE	Human glycosyl sulfotransferase-4beta (GST-4beta) cDNA.					
XX	Human; glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive;					
KW	therapy; selectin binding inhibitor; gene therapy; inflammation;					
KM	systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;					
KW	polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;					
KM	glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitalitis;					
KW	Hashimoto's disease; Grave's disease; hypoparathyroidism; anemia;					
KM	demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;					
KW	myocarditis; adult respiratory distress syndrome; eczema; psoriasis;					
KM	asthma; hypersensitivity; rheumatic fever; tissue rejection;					
KW	chromosome 16q23.1; ss.					
OS	Homo sapiens.					
XX						
FH	Key	Location/Qualifiers				
FT	5'UTR	9..188				
FT		/+tag- a				
FT	CDS	189..1376				
FT		/+tag- b				
FT		/product= "Human glycosyl sulfotransferase-4beta				

Db 99364 ACAGCTCGAGCGCTGATCCATACATACACCCAGGATCTGGATCGCGCCGCGCA 99423
 Qy 1355 TSCCTTCCTACTTAACGAGAGCGCCCTCAACGCTCTCAGCGGCTGGTGCCTT 1414
 Db 99424 AGCCTTCAGACTGTGTGAGAGATGCGCTCAACGCTCTCCAGCGCTGCGCCGCT 99483
 Qy 1415 ACCTTACGAAAGGTTTCCAGCTTCAGAGCTCGCGGTGAGGCTATGATTTGCTGG 1474
 Db 99484 GCCCTTGGCAAGATCCGCGCGGTGAGAGACTGTGCGCTGCGCTGAGCTGCTGG 99543
 Qy 1475 ATACTCCAGTCATCTCAACAGAAAGCAACCTGCTCCCTGATCTCTG 1529
 Db 99544 CTACGGCTGTCTACTGTGAGAGAGAGAGCGCAACCTCGCTGATCTGCTG 99598
 RESULT 13
 AAS16948
 AAS16948 standard; cDNA; 1937 BP.
 AAS16948;
 12-MAR-2002 (first entry)
 DE Murine intestinal-GlcNAc 6-sulfotransferase (I-GlcNAc6ST) cDNA.
 XX Mouse; betal,3gnt; betal,3-N-acetylglicosaminyltransferase; MECA-79; ss;
 XX L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
 XX ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;
 XX allergic contact dermatitis; lymphoma; chronic pneumonia; ISSF-2;
 XX delayed-type hypersensitivity reaction; hyperplastic thymus; antileuk;
 XX anti-inflammatory; antipsoriatic; antidiabetic; dermatological;
 XX antiallergic; intestinal-GlcNAc 6-sulfotransferase; I-GlcNAc6ST.
 OS Mus musculus.
 FH Key Location/Qualifiers
 FT CDS 75..1262
 FT /tag= a
 FT /product= "Murine I-GlcNAc6ST"
 PN WO200185177-A1.
 XX 15-NOV-2001.
 PD 10-MAY-2001; 2001WO-US15452.
 PF 11-MAY-2000; 2000US-0569320.
 XX (BURN-) BURNHAM INST.
 PA Fukuda M, Yeh J, Hiraoka N;
 XX WPI; 2002-075226/10.
 DR P-PSDB; AAI11275.
 XX New enzyme, useful for modifying acceptor molecule, comprises an
 PT isolated L-selectin sulfotransferase-2 that directs expression of
 PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
 XX intestinal GlcNAc 6-sulfotransferase
 PS Claim 26; Fig 10; 98pp; English.
 XX The present invention provides a method of modifying an acceptor molecule
 CC by contacting the acceptor with an isolated
 CC betal,3-N-acetylglicosaminyltransferase (betal,3gnt) or an active
 CC fragment, where betal,3gnt directs expression of a MECA-79 antigen. The
 CC invention also provides a method of treating or preventing an
 CC L-selectin-mediated condition by reducing the expression or activity of a
 CC betal,3gnt that directs expression of a MECA-79 antigen. This can be done
 CC by administering to the subject an oligosaccharide L-selectin antagonist
 CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
 CC administering antibody material that specifically binds betal,3gnt,
 CC and/or a betal,3gnt antisense nucleic acid molecule. L-selectin

CC sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.
 CC Alternatively, the expression or activity of LST-2 or its active
 CC fragment can be reduced in combination with reducing the expression or
 CC activity of betal,3gnt. The method is useful for treating L-selectin
 CC mediated conditions such as Crohn's disease and ulcerative colitis,
 CC inflammatory disorders of the skin such as allergic contact dermatitis,
 CC psoriasis and Lichen planus, lymphomas, chronic pneumonia, delayed-type
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
 CC sequence represents cDNA encoding mouse I-GlcNAc6ST.
 XX
 SQ Sequence 1937 BP; 387 A; 573 C; 543 G; 434 T; 0 other;
 Query Match 18.5%; Score 355.6; DB 24; Length 1937;
 Best Local Similarity 60.8%; Pred. No. 1.2e-91;
 Matches 618; Conservative 0; Mismatches 389; Indels 9; Gaps 2;
 Qy 520 GGAGCGCGGTGATGCTGTGCTGTCTTCGCGGCGGAGATCCTTTGTGGAGC 579
 Db 187 GGGAGCGGTGACGCTGTGTACTCTCTCGGCGCTCGGCTGCTCTTGTGGCC 246
 Qy 580 AGCTTTTCGGGAGAGACCGGATGTGTCTTACCTGATGAGCCTGCGCATGTGGA 639
 Db 247 AGCTTTTCAGCAACACCCGATGTCTTACCTGATGAGCCTGCGCATGTGGA 306
 Qy 640 TGACTTTACACAGACACACAGCTGGAAGCTGACATGCTGTGCGGATCTTCTGCT 699
 Db 307 ATACGCTTGTGCGAGGCGAGTGCCTCCGACATCCATGCGCGCTGATCCGCT 366
 Qy 700 CCGCTTCCTGCTGATGATGAGCTTGTGATGCTGATGATGATGATGATGATGATGAT 759
 Db 367 CAGTGTCTTATGCGATGAGCTGATGATGATGATGATGATGATGATGATGATGAT 423
 Qy 760 AGTCAGCCTCTTCCAGTGGAGAGCAAGGCGCGCTGTGCTGAGCGCTGTGTGACT 819
 Db 424 TCTCGAGATCTTCCAGTGGAGAGCAAGGCGCGCTGTGCTGAGCGCTGTGTGACT 483
 Qy 820 TCTTCCCTGCGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 879
 Db 484 CTTTCCCTGCGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 543
 Qy 880 CTTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 939
 Db 544 CTTTGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 603
 Qy 940 GTTTCATGAGCTGAGGCGCTTATGATGATGATGATGATGATGATGATGATGATGAT 999
 Db 604 GCTTCTTAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 663
 Qy 1000 TGTGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1059
 Db 664 TGTGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 723
 Qy 1060 AACTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1119
 Db 724 CGTGGACAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 783
 Qy 1120 ACCAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1179
 Db 784 CCGGCTGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 843
 Qy 1180 TCCAAACCTTCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1239
 Db 844 TGCACAGAGCGCGCTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 903
 Qy 1240 TTGCGGACCCCTGCGGCGGAGAGACACCACTATATATTTGTGGGTTGGATTTTTC 1299
 Db 904 CCGGAGACCCCTGCGGCGGAGAGACACCACTATATATTTGTGGGTTGGATTTTTC 963
 Qy 1300 CCACCTTCCAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1353
 Db 964 CGAGCTTCCAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1023
 Qy 1354 ATGCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1413

Db 1178 CATACCTCGTCTAGGAATGGCGCAGCTCTCCAGGCTGGCGSCACAGCGTTGCCCTTC 1237
 QY 1422 GAAAAGGTTTCCAGCTTCAAGATGCTGGGGTGAGGCTAATGATTTGCTGGATACCTC 1481
 Db 1238 ACTAAGATCCTGCGCGTGCAGAGAGGTGTGGCGCGCGCTGCAGTGTGGGCTACCGG 1297
 QY 1482 CAGGTCAAGTCTCAACAAGAACAAGCAACCTGTCCCTGGATCTTCTG 1529
 Db 1298 CCTGTGTACTGTGGCGACAGCAGCTGACCTCACCTGGATCTGTG 1345

Search completed: August 14, 2003, 11:38:24
 Job time : 482.467 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 11:17:45 ; Search time 3643.34 Seconds
(without alignments)
12848.224 Million cell updates/sec

Title: US-09-645-078-3
Perfect score: 1926
Sequence: 1 gggcactcaactcaactg.....ccgtcacatcttctaag 1926

Scoring table: IDENTITY_NDC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estda:*
2: em_estdm:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrt:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1499.8	77.9	1923	11	AK009113 Mus muscu
2	804.8	41.8	902	13	BQ947021 AGENCOURT
3	635.6	33.0	783	12	BG963298 602827716
4	628.4	32.6	965	10	BF579746 602095056

5	604.4	31.4	695	12	BG964671
6	602.2	31.3	852	12	BG966340
7	559	29.0	571	12	BM245312
8	548.8	28.5	656	9	AA522184
9	449.5	25.7	507	12	BM246681
10	434.2	22.5	954	12	BI823850
11	353.2	18.3	803	14	BY708952
12	346	18.0	668	9	AL709927
13	305	15.8	583	2	BSM082755
14	292.4	15.2	571	12	BM129080
15	267	13.9	834	13	BQ884877
16	249	12.9	804	13	BU135348
17	247	12.8	695	10	BE857538
18	245.8	12.8	295	9	A1614882
19	238.2	12.4	692	14	BY711808
20	234.4	12.2	735	10	BE858652
21	228.4	11.9	620	9	A1824100
22	225.8	11.7	904	13	BU241318
23	221.4	11.5	773	13	BU458165
24	217.4	11.3	703	14	CD218830
25	213.6	11.1	1067	29	CNS03KN7
26	210.6	10.9	768	13	BU342120
27	202.2	10.5	692	13	BU256241
28	195.4	10.1	525	10	BF197521
29	187.2	9.7	516	9	AA689056
30	187	9.7	527	4	BX519156
31	167	8.7	849	29	CNS04OEN
32	165.8	8.6	568	9	A1115260
33	159.4	8.3	543	13	BU382276
34	153.8	8.0	555	4	BX514916
35	151	7.8	517	10	BF042384
36	148.6	7.7	657	29	AG035205
37	147.6	7.7	433	28	AZ405100
38	135.8	7.1	429	10	BB849113
39	131.8	6.8	758	13	BU259678
40	131.6	6.8	599	12	BU031352
41	129.8	6.7	436	9	AM081348
42	126.2	6.6	498	13	BX280599
43	124	6.4	830	14	CB558413
44	122.6	6.4	961	13	BQ734733
45	121.2	6.3	423	13	BY237139

ALIGNMENTS

RESULT 1
AK009113
LOCUS
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310003618 product:carbohydrate (chondroitin 6/keratan) sulfotransferase 4, full insert sequence.
ACCESSION AK009113
VERSION AK009113.1 GI:12843701
KEYWORDS
SOURCE HTC; CAP trapper.
ORGANISM Mus musculus (house mouse)
REFERENCE
1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
JOURNAL
MEDLINE
PUBMED
11042159

1923 bp mRNA linear HTC 05-DEC-2002
Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310003618 product:carbohydrate (chondroitin 6/keratan) sulfotransferase 4, full insert sequence.
AK009113
AK009113.1 GI:12843701
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159


```

Db      572 CTTGATATGTTGAGAGAGCCCTGCTCAGAGGCTGCTGCTACTCAGAGAGTGCG 631
QY      941 TTTTCTCAGCGTGGAGGCGCCCTATCAGTACTCAGAGGACCCCTCCCAACGCGACGT 1000
Db      632 TTTTCTCAGCGTGGAGGCGCCCTATCAGTACTCAGAGGACCCCTCCCAACGCGACGT 691
QY      1001 CGTGACACCTGCTGAGAGAGCCCGGGCGTGTCCGATCCCGGAGACACCAACATAGA 1060
Db      692 CGTGACACCTGCTGAGAGAGCCCGGGCGTGTCCGATCCCGGAGACACCAACATAGA 751
QY      1061 ACTCATGTTGACAGTATATTTGCTAGGGGACCACTTTGGAAACATCAAGAGAGAGA 1120
Db      752 ACTCATGTTGACAGTATATTTGCTAGGGGACCACTTTGGAAACATCAAGAGAGAGA 811
QY      1121 CGAGCCCTATTATGCGCATGCAAGATCATCTGCAAAAAGCAGAGTGACATAGCAGGCCAT 1180
Db      812 CGAGCCCTATTATGCGCATGCAAGATCATCTGCAAAAAGCAGAGTGACATAGCAGGCCAT 871
QY      1181 CCAAAACCTCCCTGAAAGCTGTCAGACAGCCCTACTCTCTCTAGATATGAGACCTGCT 1240
Db      872 CCAAAACCTCCCTGAAAGCTGTCAGACAGCCCTACTCTCTCTAGATATGAGACCTGCT 931
QY      1241 TCGGGGACCCCTGGCCGAGACGACAGACATATATAATTTGTGGGTTGATTTTGGC 1300
Db      932 TCGGGGACCCCTGGCCGAGACGACAGACATATATAATTTGTGGGTTGATTTTGGC 991
QY      1301 CGACCTCCAAACATGAGTTTACATATGTCACCCCGGGGGAAGGAGTGGGTACAGTGCCTT 1360
Db      992 CGACCTCCAAACATGAGTTTACATATGTCACCCCGGGGGAAGGAGTGGGTACAGTGCCTT 1051
QY      1361 CCATTAACAGCAGAGAGCCCTCAACGCTCTCTCAGGCGTGGCGTGGTCTTACCTTA 1420
Db      1052 CCATTAACAGCAGAGAGCCCTCAACGCTCTCTCAGGCGTGGCGTGGTCTTACCTTA 1111
QY      1421 CGAAAGGTTTCCAGCTTCAAGATGCTGCGGTGAGGCTATGATTTGCTGGGATACCT 1480
Db      1112 CGAAAGGTTTCCAGCTTCAAGATGCTGCGGTGAGGCTATGATTTGCTGGGATACCT 1171
QY      1481 CGAGGTGATTCACAAAGAGAGCAAGCAACGCTGCTCCGATGCTGCTGCTCCCA 1540
Db      1172 CGAGGTGATTCACAAAGAGAGCAAGCAACGCTGCTCCGATGCTGCTGCTCCCA 1231
QY      1541 TATCTGGGGGAGGCTTTCGAGAGAGTTAAGAGAGTCTGCTGTCACCCCTGGTTCAG 1600
Db      1232 TATCTGGGGGAGGCTTTCGAGAGAGTTAAGAGAGTCTGCTGTCACCCCTGGTTCAG 1291
QY      1601 CCTTAGTACCAATTTAAAGCAGAGAGCCCTTAAGTATAACCAACTGAGTGCCTTTC 1660
Db      1292 CCTTAGTACCAATTTAAAGCAGAGAGCCCTTAAGTATAACCAACTGAGTGCCTTTC 1351
QY      1661 TCCGACAGCCCAAGCAGAGAGGCTTGTGCTATATCTCATGCTGCTTACCTCAACATAGC 1720
Db      1352 TCCGACAGCCCAAGCAGAGAGGCTTGTGCTATATCTCATGCTGCTTACCTCAACATAGC 1411
QY      1721 CTAAGAAAGCAAGAAAGATATCTTCTGCTTGAATAATAGTAAGAACCTTAAGCAGCC 1780
Db      1412 CTAAGAAAGCAAGAAAGATATCTTCTGCTTGAATAATAGTAAGAACCTTAAGCAGCC 1471
QY      1781 CCTTAGACCTGTCAGAGAAAGATTTCTTGAACCTTGGCTTCTTACCTGTGATACCTT 1840
Db      1472 CCTTAGACCTGTCAGAGAAAGATTTCTTGAACCTTGGCTTCTTACCTGTGATACCTT 1521
QY      1841 GAGAGACT-CGGTGTGAGAGCATACTGAGACAGAGAAACAGATGTGAGAGTGATGT 1899
Db      1532 GAGAGACTGAGGCTGTGAGAGCATACTGAGACAGAGAAACAGATGTGAGAGTGATGT 1591
QY      1900 AAAACCTCCTGTACATCTTTTC 1922
Db      1592 AAAACCTCCTGTACATCTTTTC 1614

```

```

B0947021
LOCUS      B0947021          902 bp    mRNA    linear    EST 21-AUG-2002
DEFINITION AGENCOURT 8951359 NCI CGAP Co24 Mus musculus cDNA clone
IMAGE:6474928 5', mRNA sequence.
ACCESSION  B0947021
VERSION    B0947021.1 GI:22362499
KEYWORDS
SOURCE     Mus musculus (house mouse)
ORGANISM  Mus musculus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1 (bases 1 to 902)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LMNL at:
            http://image.llnl.gov
            Plate: LM414012 row: n column: 17
            High quality sequence stop: 633.
FEATURES
     source
         location/qualifiers
             1..902
             /organism="Mus musculus"
             /mol_type="mRNA"
             /strain="FVB/N"
             /db_xref="taxon:10090"
             /clone="IMAGE:6474928"
             /lab_host="DH10B (T1 phage-resistant)"
             /note="Organ: colon; Vector: pCMV-Sport6; site_1: NotI;
             site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
             Average insert size 1.6 kb. Constructed by Life
             Technologies. Note: this is a NCI CGAP library."
BASE COUNT  228 a 248 c 210 g 213 t 3 others
ORIGIN
Query Match      41.8%; Score 804.8; DB 13; Length 902;
Best Local Similarity 97.4%; Pred. No. 1.5e-136;
Matches 827; Conservative 0; Mismatches 20; Indels 2; Gaps 1;
QY      1048 ACACCACCATGAACTCATGTTGACAGTCAATTTGCTAGAGGCGACGATTTGAAACGA 1107
Db      1 ACACCACCATGAACTCATGTTGACAGTCAATTTGCTAGAGGCGACGATTTGAAACGA 60
QY      1108 TCAAGAGAGAGAGACAGCCCTATATGCGATGAAAGATCATGTGCAAAAAGCAGGTGGACA 1167
Db      61 TCAAGAGAGAGAGACAGCCCTATATGCGATGAAAGATCATGTGCAAAAAGCAGGTGGACA 120
QY      1168 TAGTCAAGGCCATCAAAACCCCTCCCTGAAGCTCTGACAGCAGCGTACTGTTCTGTAGGT 1227
Db      121 TAGTCAAGGCCATCAAAACCCCTCCCTGAAGCTCTGACAGCAGCGTACTGTTCTGTAGGT 180
QY      1228 ATGAGAGACTGTGTTGGGGGACCCCTGGCCGACAGACAGACTATATTAATTTGTGGGGT 1287
Db      181 ATGAGAGACTGTGTTGGGGGACCCCTGGCCGACAGACAGACTATATTAATTTGTGGGGT 240
QY      1288 TGAATTTTGGCCCACTCCCAACATGAGGTTTCAATGTACCCGGGCAAGGCGATGG 1347
Db      241 TGAATTTTGGCCCACTCCCAACATGAGGTTTCAATGTACCCGGGCAAGGCGATGG 300
QY      1348 GTGAGCATGCTTCATTAAGCCAGAGAAAGCCCTCAAGCTCTCAGGCTGGCGTT 1407
Db      301 GTGAGCATGCTTCATTAAGCCAGAGAAAGCCCTCAAGCTCTCAGGCTGGCGTT 360
QY      1408 GGTCCCTACCTTAGAGAAAGGTTTCCAGAGCTTCAAGATGCTGCGGTGAGAGCTATGATTT 1467
Db      361 GGTCCCTACCTTAGAGAAAGGTTTCCAGAGCTTCAAGATGCTGCGGTGAGAGCTATGATTT 420

```

QY 1468 TECTGGATACCTCCAGCTCAGATCTCAACAAAGAACGCAACCTGTCCTGATCTTC 1527
 DB 421 TECTGGATACCTCCAGCTCAGATCTCAACAAAGAACGCAACCTGTCCTGATCTTC 480
 QY 1528 TGTCTCCCTCCATATCTTGGGGGAGGTCTCCGAGAAAGTAAAGAGTCTGTCTGCAC 1587
 DB 481 TGTCTCCCTCCATATCTTGGGGGAGGTCTCCGAGAAAGTAAAGAGTCTGTCTGCAC 540
 QY 1588 CCCTTGCTCCAGCTTGTAGTCACTTAAACGACAGAGCTTAAAGTATTAACCAACT 1647
 DB 541 CCCTTGCTCCAGCTTGTAGTCACTTAAACGACAGAGCTTAAAGTATTAACCAACT 600
 QY 1648 GAGTGGCCCTTCTCTCTAGCCCAAGCAGAGGGCTTGTGTCTATCTATCTCTAC 1707
 DB 601 GAGTGGCCCTTCTCTCTAGCCCAAGCAGAGGGCTTGTGTCTATCTATCTAGCTCTAC 660
 QY 1708 CCTAACCTGAGCTTAAAGCAAGAAACAGTATCTTCTGCTTGAATATCTTAGA 1767
 DB 661 CCTAACCTGAGCTTAAAGCAAGAAACAGTATCTTCTGCTTGAATATCTTAGA 720
 QY 1768 ACCTTAAGCAGCCCTTGTAGCTTCAAGCAAGACTTCTTAACTTGGCCTTCTAC 1827
 DB 721 ACCTTAAGCAGCCCTTGTAGCTTCAAGCAAGACTTCTTAACTTGGCCTTCTAC 780
 QY 1828 CTGTGCATACCTTGAAGACT--CGTCTGAGGACATCTGACACAGCAAAAGACTCTG 1885
 DB 781 CTGTGCATACCTTGAAGACTTGAAGCTGTGAGGCCCTACTGTGACACAAACANTCA 840
 QY 1886 TGGAGTGTG 1894
 DB 841 TGGAGTGTG 849
 RESULT 3
 LOCUS BG963298 783 bp mRNA linear EST 12-JUN-2001
 DEFINITION 602827176F1 NCI_GAP_Co24 Mus musculus cDNA clone IMAGE:4982643 5',
 mRNA sequence.
 ACCESSION BG963298.1 GI:14350935
 VERSION BG963298.1
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 783)
 NIH-MGC http://mgc.ncl.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 plate: ILAM10985 row: P column: 04
 High quality sequence stop: 741.
 Location/Qualifiers
 1..783
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4982643"
 /lab_host="DH10B (TI phage-resistant)"
 /clone_lib="NCI_GAP_Co24"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site:1; NotI;
 Site:2; SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.6 kb. Constructed by Life

BASE COUNT 141 a 236 c 226 g 180 t
 ORIGIN
 Query Match 33.0%; Score 635.6; DB 12: Length 783;
 Best Local Similarity 98.0%; Pred. No. 9.6e-106;
 Matches 656; Conservative 0; Mismatches 9; Indels 5; Gaps 5;
 QY 398 CCGCAGATGATCTGTGAAGAAAGAGGCTGTGATGTCTCTGGTCCAGTCAAT 457
 DB 60 CCTAAGATGATGCTGTGAAGAAAGAGGCTGTGATGTCTCTGGTCCAGTCAAT 119
 QY 458 CGTGTAGCTCTCTTATCATATATCTCCACAGACACCTTCCCAAGAGAGAGTC 517
 DB 120 CGTGTAGCTCTCTTATCATATATCTCCACAGACACCTTCCCAAGAGAGAGTC 179
 QY 518 CAGAGAGCCGTCATATGCTGTGTCTCTCTGCGCGTGAAGATCTCTTTTGGG 577
 DB 180 CAGAGAGCCGTCATATGCTGTGTCTCTCTGCGCGTGAAGATCTCTTTTGGG 239
 QY 578 ACAGCTTTTGGGACACACCCGAGATGTCTTACTATGATGAGACCTGCTGGCATGTG 637
 DB 240 ACAGCTTTTGGGACACACCCGAGATGTCTTACTATGATGAGACCTGCTGGCATGTG 299
 QY 638 GATGACTTTCACACAGACACAGCCGGAAGCTGCACATGGCTGTGCGGAGTCTTCTGG 697
 DB 300 GATGACTTTCACACAGACACAGCCGGAAGCTGCACATGGCTGTGCGGAGTCTTCTGG 359
 QY 698 TTCCGCTCTCTGTGACATGAGCTCTTGTGATGCTTACATGAAACCAAGGCCCGGAA 757
 DB 360 TTCCGCTCTCTGTGACATGAGCTCTTGTGATGCTTACATGAAACCAAGGCCCGGAA 419
 QY 758 ACAGTCCAGCTCTTCCATGAGGAGCAAAAGCCGCGCTGTGCTACAGCCCTGTGTGA 817
 DB 420 ACAGTCCAGCTCTTCCATGAGGAGCAAAAGCCGCGCTGTGCTACAGCCCTGTGTGA 479
 QY 818 CTTCCTCCCTGGCCACAGATCAGCTCACCACAGCTGCAAGCTCTCTGGGGTGA 877
 DB 480 CTTCCTCCCTGGCCACAGATCAGCTCACCACAGCTGCAAGCTCTCTGGGGTGA 539
 QY 878 GCCCTTGTATGATGAGGAGAGCCCTGCGCTCTCAAGGCTTGTGTACTCA--GAGG 936
 DB 540 GCCCTTGTATGATGAGGAGAGCCCTGCGCTCTCAAGGCTTGTGTACTCAAGGAG 599
 QY 937 TGGGT-TTTCCTAGCTCTGAGGCTCTTATCCACTACTACGAGACCTTCCCTCAACTG 995
 DB 600 TGGGTATTCCTAGCTCTGAGGCTCTTATCCACTACTACGAGACCTTCCCTCAACTG 659
 QY 996 -CAGCTGCGACACCTGTCGAGAGACCCGCGCGCTTCCGATCCCGGAGACACAC 1054
 DB 660 CCACGTGTCACACTGTGTCGAGA--CCCGGCGCGCTTCCGATCCCGGAGACACAC 718
 QY 1055 CATGACACATGATGTGA--CAGTCATATGTGTCTAGGAGCATTTGAA 1103
 DB 719 CATGACACATGATGTGA--CAGTCATATGTGTCTAGGAGCATTTGAA 768
 RESULT 4
 LOCUS BF579746 965 bp mRNA linear EST 12-DEC-2000
 DEFINITION 602095056F1 NCI_GAP_Co24 Mus musculus cDNA clone IMAGE:4209636 5',
 mRNA sequence.
 ACCESSION BF579746
 VERSION BF579746
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 965)
 NIH-MGC http://mgc.ncl.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov
Plate: LLM9775 row: 0 column: 13
High quality sequence stop: 663.
Location/Qualifiers

FEATURES
source

1. 965
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4209636"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Co24"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI CGAP library."
BASE COUNT 233 a 300 c 243 g 189 t
ORIGIN

Query Match 32.6%; Score 628.4; DB 10; Length 965;
Best Local Similarity 95.4%; Pred. No. 1.9e-104; Indels 10; Gaps 9;
Matches 744; Conservative 0; Mismatches 26;

QY 899 GGCGTGGCGCTCTCAGCGCTTCGTGACTACAGAGAGTGGTTTCTCAGCCGACAGC 958
DB 2 GGCGTGGCGCTCTCAGCGCTTCGTGACTACAGAGAGTGGTTTCTCAGCCGACAGC 61
QY 959 CCTTATTCACCTACTCAGAGACCTTCCCTCAGCTGACGCTGACCTGACCTGAC 1018
DB 62 CCTTATTCACCTACTCAGAGACCTTCCCTCAGCTGACGCTGACCTGACCTGAC 121
QY 1019 CCCCCGGGCGGTGTCGATCCGGGAGACACACCATTAACATCATGTTGACAGTCA 1078
DB 122 CCCCCGGGCGGTGTCGATCCGGGAGACACACCATTAACATCATGTTGACAGTCA 181
QY 1079 TATGTGCTAGGGAGCATTTGGAACGATCAGAGAGAGACAGCCCTATTATGACAT 1138
DB 182 TA-TGTGCTAGGGAGCA-TTGGAAAGCATCAGAGAGAGACAGCCCTATTATGACAT 239
QY 1139 GAATATCATCTGCAAAAGCCAGGTGACATAGTCAAGGCCATCCAAACCTCCCTGAA 1198
DB 240 GAATATCATCTGCAAAAGCCAGGTGACATAGTCAAGGCCATCCAAACCTCCCTGAA 299
QY 1199 TCTGACAGAGCGTACCTGTTCTGAGGTATGAGGACCTGTTGGGGACCCCTGGCCCA 1258
DB 300 TCTGACAGAGCGTACCTGTTCTGAGGTATGAGGACCTGTTGGGGACCCCTGGCCCA 359
QY 1259 GAGCAGCAGACTATATAATTGTTGGGGTGTGATTTTGGCCCACTCCAAAGATGGGT 1318
DB 360 GAGCAGCAGACTATATAATA-TTGTGGGGTGTGATTTTGGCCCACTCCAAAGATGGGT 418
QY 1319 TTAAATATGTCACCCGGGCAAGGGATGGTACAGATGCTTCATTAAGCCAGGAA 1378
DB 419 TCACAAATGTCACCCGGGCAAGGGATGGTACAGATGCTTCATTAAGCCAGGAA 478
QY 1379 CGGCTCAAGCTCTCTAGGCGGTGGCGTTGCTCTACCTTA-CGAAAAGTTTCCAGC 1437
DB 479 CGGCTCAAGCTCTCTAGGCGGTGGCGTTGCTCTACCTTAAGCGAAAAGTTTCCAGC 538
QY 1438 TTCAAGATGCTGCGGTGAGGCTATGGA-TTGTGTTGAGTACTCTCAGAGTCAAGTCTCA 1496
DB 539 TTCAAGATGCTGCGGTGAGGCTATGGA-TTGTGTTGAGTACTCTCAGAGTCAAGTCTCA 598
QY 1497 CAAGACAAAGCAACCTGCTCCCTGATCTTCTGCTCTCCCATATCTGGGGCAGGTC 1556
DB 1556 CAAGACAAAGCAACCTGCTCCCTGATCTTCTGCTCTCCCATATCTGGGGCAGGTC

DB 599 CAAGACAAAGCAA-CTGTCCCTGATCTCTGTCCTCCATCATCTGGGGCAGGTC 657

QY 1557 TTCGAGAAAGTTAAGAGGTCTGTCTCACCCTTTGGTTCCAGGC--TTAGTACCAATT 1614

DB 658 TTCGAGAGCAGGTTAAGAGGCTG-CTCCACCCCTGGTTCCAGGCTTAAGCCACCAATT 716

QY 1615 AAAGCAGAGAGGCTTAAGGATATACCAAACTGATGCCCTTTCTCAGAGCCCAAG 1674

DB 717 TAAGCAGAGGAG-CTTAAGTAGAACCAACTGATGAGTCCCTTTCTTACAGCCAGCAG 775

RESULT 5
BG964671 695 bp mRNA Linear EST 12-JUN-2001
LOCUS 602831875P1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:498258 5',
DEFINITION mRNA sequence.

ACCESSION BG964671 GI:14352308

VERSION BG964671.1

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov
Plate: LLM10995 row: f column: 19
High quality sequence stop: 684.
Location/Qualifiers

1. 695
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:498258"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Co24"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI CGAP library."
BASE COUNT 118 a 216 c 195 g 166 t
ORIGIN

Query Match 31.4%; Score 604.4; DB 12; Length 695;
Best Local Similarity 98.4%; Pred. No. 4.7e-100; Indels 5; Gaps 5;
Matches 663; Conservative 0; Mismatches 6;

QY 351 AAGCTCTCTCTTCTTCCGAGAGATGATGCTGTGAAGAAAGGAGGCTGTATGTTTC 440
DB 12 AAGCTCTCTCTCTTCTTCCGAGAGATGATGCTGTGAAGAAAGGAGGCTGTATGTTTC 71
QY 441 CTGGGTCCAGAGTATGTTGATCTCTCTCATCCATATGTCGT-CCAGACAGACT 499
DB 72 CTGGGTCCAGAGTATGTTGATCTCTCTCATCCATATGTCGTCCACAGACACT 131
QY 500 TTCCAGAGAGAGAGTCCAGAGAGCCCGTGAATGCTGTGCTCTTCTGCGGCTC 559
DB 132 TTCCAGAGAGAGAGTCCAGAGAGCCCGTGAATGCTGTGCTCTTCTGCGGCTC 191
QY 560 AGATCTCTTTTGTGGGACAGCTTTTGGGACAGACCCGAGTGTCTTCACTGATGGA 619
DB 192 AGATCTCTT-TTTGTGGGACAGCTTTTGGGACAGACCCGAGTGTCTTCACTGATGGA 250

QY 620 GCGTGGCTGGATGTGTGATGACTTTACACAGACAGACAGCCTGGAGCTGACATGAC 679
 DB 251 GCGTGGCTGGATGTGTGATGACTTTACACAGACAGACAGCCTGGAGCTGACATGAC 310
 QY 680 TGTGGGGGATCTTGTGCGCTTCCGTCTCTCTGTGTGTACATGAGCGCTTTGATGCTTACAT 739
 DB 311 TGTGGGGGATCTTGTGCGCTTCCGTCTCTGTGTGTACATGAGCGCTTTGATGCTTACAT 370
 QY 740 GAACCCAGGCCCCGGAAACAGTCCAGCCTTCTCCAGTGGAGCAAAAGCCGGCCCTGTG 799
 DB 371 GAACCCAGGCCCCGGAAACAGTCCAGCCTTCTCCAGTGGAGCAAAAGCCGGCCCTGTG 430
 QY 800 CTCAGCGCTGTGTGTGACTTCTTCCCTGCCACGAGATCAGCTCACCAGACATGCAA 859
 DB 431 CTCAGCGCTGTGTGTGACTTCTTCCCTGCCACGAGATCAGCTCACCAGACATGCAA 490
 QY 860 GGTGCTGCGGTGTCAGACAGCCTTGTATATGTGTGAGGAAGCGCTGCCCTCTCAGCGCTT 919
 DB 491 GGTGCTGCGGTGTCAGACAGCCTTGTATATGTGTGAGGAAGCGCTGCCCTCTCAGCGCTT 550
 QY 920 CGTGTACTCAAGAGAGTGCCTTCTCTAGCCTGACAGCCCTCTATCCACTACTACAGGA 979
 DB 551 CGTGTACTCAAGAGAGTGCCTTCTCTAGCCTGACAGCCCTCTATCCACTACTACTACAGGA 610
 QY 980 CCCTT-CCCTCAACCTGACAGT-CGTGACACTGTGTCCAGAGACCCCGGGCCGTGTCCGA 1037
 DB 611 CCCTTCCCTCAACCTGACAGTCCGTGACACTGTGTCCAGAG-CCCGGGCCGTGTCCGA 669
 QY 1038 TCCCGGAGACACAC 1051
 DB 670 TCCCGGAGAGCCAC 683

RESULT 6
 BG966340 852 bp mRNA 1linear EST 12-JUN-2001
 LOCUS 602832826F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4987327 5'
 DEFINITION mRNA sequence.
 ACCESSION BG966340
 VERSION BG966340.1 GI:14353977
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 852)
 NIH-MGC http://mgc.ncl.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgrabs-r@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LINL)
 DNA Sequencing by: Inqyle Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
 http://image.llnl.gov
 Plate: L14M10998 row: c column: 08
 High quality sequence stop: 784.
 Location/Qualifiers
 1..852

FEATURES
 source
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4987327"
 /lab_host="DH10B (TI phage-resistant)"
 /clone_lib="NCI_CGAP_Co24"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site1: NotI;
 Site2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.6 kb. Constructed by Life

BASE COUNT 176 a 248 c 244 g 184 t
 ORIGIN
 Query Match 31.3%; Score 602.2; DB 12: Length 852;
 Best Local Similarity 96.4%; Pred. No. 1.1e-99;
 Matches 670; Conservative 0; Mismatches 18; Indels 7; Gaps 5;

QY 401 CAGATGATGCTGTTGAAGAAAGGAGGCTGCGATGTTCTCTGAGTCCAGATCATGCT 460
 DB 133 CAGATGATGCTGTTGAAGAAAGGAGGCTGCGATGTTCTCTGAGTCCAGATCATGCT 192
 QY 461 TGTAGCTCTCTCATCATATGTCGTCCACAGACACCTTTCCAGAGGAGAGTCCAG 520
 DB 193 TGTAGCTCTCTCATCATATGTCGTCCACAGACACCTTTCCAGAGGAGAGTCCAG 252
 QY 521 GAGCGCCGTCATGTCGTGTCGTCTCTGTCGCGGCGGTGAGATCCTCT-TTTGTGGAG 579
 DB 253 GAGCGCCGTCATGTCGTGTCGTCTCTGTCGCGGCGGTGAGATCCTCTATATGTGGAG 312
 QY 580 AGCTTTCCGGCAGCACCAGATGTTCTTACCTGATGAGCCTGCCATGTTGGA 639
 DB 313 AGCTTTCCGGCAGCACCAGATGTTCTTACCTGATGAGCCTGCCATGTTGGA 372
 QY 640 TGACTTTCAACAGACAGACAGCCTGGAAGCTGCATATGCTGTGGATCTTTCGCTT 699
 DB 373 TGACTTTCAACAGACAGACAGCCTGGAAGCTGCATATGCTGTGGATCTTTCGCTT 432
 QY 700 CCGTCTCTGTCGTGTCATGAGAGGCTTGTGATGCTTCAATGAAACCCAGGCCCGGAAAC 759
 DB 433 CCGTCTCTGTCGTGTCATGAGAGGCTTGTGATGCTTCAATGAAACCCAGGCCCGGAAAC 492
 QY 760 AGTCAGCCTCTTCCAGTGGAGCAAGCCGGCCCTGTGTCAGAGCCTGTGTGACT 819
 DB 493 AGTCAGCCTCTTCCAGTGGAGCAAGCCGGCCCTGTGTCAGAGCCTGTGTGACT 552
 QY 820 TCTTCCCTGCCACAGAGATCAGCTACCCAGACATGCAAGCTCTCTCGGTGACAGC 879
 DB 553 TCTTCCCTGCCACAGAGATCAGCTACCCAGACATGCAAGCTCTCTCGGTGACAGC 612
 QY 880 CCTTGTATATGTTGGAAGAGCGCTGCCGTCTCAGAGCTGTGGTACTCAAGAGAGTGC 939
 DB 613 CCTTGTATATGTTGGAAGAGCGCTGCCGTCTCAGAGCTGTGGTACTCAAGAGAGTGC 672
 QY 940 GTTTCCTCAGCCTGACAGGCTCTTATCCACTACTACAGAGCCT--TCCCTCAACCTG--C 996
 DB 673 GTATTCCTCAGCCTGACAGGCTCTTATCCACTACTACAGAGCCTTATCCCTCAACCTG 732
 QY 997 AGCTGTGCA-CCTGCTCCGAGACCCCGGCGGTGTCGATCCCGGAGACACACACC 1055
 DB 733 AGCTGTGCA-CCTGCTCCGAGACCCCGGCGGTGTCGATCCCGGAGACACACACC 792
 QY 1056 ATGAGAACATGATGTTGACAGCATATATGTTAGG 1090
 DB 793 AATGATGCA--GGTGACCAAGATATATGTTAGG 825

RESULT 7
 BM245312 571 bp mRNA 1linear EST 07-JUN-2003
 LOCUS K0722H07-3 NIA Mouse Hematopoietic Stem Cell (Lin-/c-Kit-/Sca-1-)
 DEFINITION cDNA Library (Long) Mus musculus cDNA clone NIA:K0722H07
 IMAGE:30075738 3', mRNA sequence.
 ACCESSION BM245312
 VERSION BM245312.2 GI:31500552
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 571)
 Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
 Construction of long-transcript enriched cDNA libraries from

REFERENCE
 AUTHORS
 TITLE

Query Match	Similarity	29.0%	Score 559	DB 12	Length 571
Best Local	Similarity	99.8%	Pred. No. 8.8e-92		
Matches	570	Conservative	0	Mismatches	0
				Indels	1
				Gaps	1
QY	275	GTTCGTTGTTGTTTACATTCGCTCCCTGGATGGAAATCCA-GAGAAAGCCGAAGGAG	333		
Db	1	GTTCGTTGTTGTTTACATTCGCTCCCTGGATGGAAATCCAGAGAAAGCCGAAGGAG	60		
QY	334	ATGCTGTACACCTAACTCAGCCCATCCCTCTGCTCTCTTCAAGGTCCTTCCT	393		
Db	61	ATGCTGTACACCTAACTCAGCCCATCCCTCTGCTCTCTTCAAGGTCCTTCCT	120		
QY	394	TCTTCGAGAGATATGCTGTGAAGAAGGAGGCTGTATTTCTCGGTTCCAGG	453		
Db	121	TCTTCGAGAGATATGCTGTGAAGAAGGAGGCTGTATTTCTCGGTTCCAGG	180		

```

/Organism="Mus musculus"
/mol_type="RNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:851801"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/clone_id="Soares_mammary_gland_NbMWG

```


Db 121 TCTTCGCGAGGATGATGCTGTTGAGAGAGAGGAGCTGATGTTCTGCTGCTCCAGG 180
 QY 454 TCATCGTTGAGTCTCTTCATCATATGCGTCCACAGACACCTTCCAGAGGAGG 513
 Db 181 TCATCTGTGAGTCTCTTCATCATATGCGTCCACAGACACCTTCCAGAGGAGG 240
 QY 514 AGTCAGAGAGGCGCGGATGCTGCTGCTCTCTCTGCGGTCAGAGATCCCTTTTG 573
 Db 241 AGTCAGAGAGGCGCGGATGCTGCTGCTCTCTCTGCGGTCAGAGATCCCTTTTG 300
 QY 574 TGGGAGAGCTTTTGGGAGACACCGGATGCTTCTACATGAGAGCCTGCTGGCAG 633
 Db 301 TGGGAGAGCTTTTGGGAGACACCGGATGCTTCTACATGAGAGCCTGCTGGCAG 360
 QY 634 TGGGAGAGCTTTTGGGAGACACCGGATGCTTCTACATGAGAGCCTGCTGGCAG 693
 Db 361 TGGGAGAGCTTTTGGGAGACACCGGATGCTTCTACATGAGAGCCTGCTGGCAG 420
 QY 694 TGGGAGAGCTTTTGGGAGACACCGGATGCTTCTACATGAGAGCCTGCTGGCAG 753
 Db 421 TGGGAGAGCTTTTGGGAGACACCGGATGCTTCTACATGAGAGCCTGCTGGCAG 480
 QY 754 GGAAGAGCTCCAGCTCTTCCAGTGGG 780
 Db 481 GGAAGAGCTCCAGCTCTTCCAGTGGG 507

RESULT 10 954 bp mRNA linear EST 04-OCT-2001
 LOCUS B1823850 603039012F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5179826 5',
 DEFINITION mRNA sequence.
 ACCESSION B1823850
 VERSION B1823850.1 GI:15935400
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NIH_MGC http://mgc.ncl.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
 http://image.llnl.gov
 Plate: L14M11448 row: h column: 03
 High quality sequence stop: 856.
 Location/Qualifiers
 1. 954
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5179826"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_115"
 /note="Organ: pooled brain, lung, testis; Vector:
 pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
 source anonymous pool of 6 male brains, age range 23-27; 1
 male lung, age 27; and 1 male testis, age 69. Library is
 oligo-dT primed and directionally cloned (EcoRV site is
 destroyed upon cloning). Average insert size 1.8 kb,
 insert size range 1-3 kb. Library is normalized and
 enriched for full-length clones and was constructed by C.
 Gruber (Invitrogen). Research Genetics tracking code
 021. Note: this is a NIH_MGC Library."

FEATURES
 source
 BASE COUNT 199 a 297 c 250 g 208 t

ORIGIN
 Query Match 22.5%; Score 434.2; DB 12; Length 954;
 Best Local Similarity 73.9%; Pred. No. 4.2e-69;
 Matches 660; Conservative 0; Mismatches 218; Indels 15; Gaps 8;

QY 388 TCTTCCTTCCCGAGGATGATGCTGTTGAGAGAGGAGGCTGATGTTCTGCTGCTCCAGG 447
 Db 61 TCTTCAGCTTACAGACAAATGCTACTGCTTAAAAAATGAAAGCTCGTGTCTGTTT 120
 QY 448 CCGAGGATGATGTTGAGTCTCTCTTCAATATGTCGTCACAGACACCTTCCAG- 506
 Db 121 CCGAGGATGATGTTGAGTCTCTCTTCAATATGTCGTCACAGACACCTTCCAG- 180
 QY 507 --AGGAGAGAGTCCAGAGAGGCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 564
 Db 181 CTATGAAAGGACAGCCCGGAGCGATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
 QY 565 CTTCTTTTGGGAGACACCTTTTGGGAGACCGGATGCTTCTACATGAGAGCCTG 624
 Db 241 CTTCTTTTGGGAGACACCTTTTGGGAGACCGGATGCTTCTACATGAGAGCCTG 300
 QY 625 CTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 684
 Db 301 CTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
 QY 685 GGGATCTTCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 744
 Db 361 GGGATCTTCATGAGGCGCGCTTCTTGTGACATGAGAGCCTTGTGATGCTTACATGAG 420
 QY 745 CAGGAGGCGCGGAGAGAGTCCAGGCTTCTTCCAGTGGAGAGAGAGGCGGCTTGTGCTG 804
 Db 421 CTTGCTCCCGGAGAGAGTCCAGGCTTCTTCCAGTGGAGAGAGAGGCGGCTTGTGCTG 480
 QY 805 CCGCTGTGTGATCTTCTTCCCGCAGAGATCAGCTCAGCCAGCTCAGCTCAGCTCAG 864
 Db 481 CACCTGCTGTGATCTTCTTCCCGCAGAGATGAGATGAGATGAGATGAGATGAGATGAG 540
 QY 865 TCGGCGGTGAGAGCGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 924
 Db 541 TGTGAGATGAGAGCGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 600
 QY 925 TACTCAAGAGAGTGTGCTTCTGAGCTGTGAGAGCGCTTGTGATGATGATGATGATG 984
 Db 601 TGTGAGAGAGAGTGTGCTTCTGAGCTGTGAGAGCGCTTGTGATGATGATGATGATG 660
 QY 985 CCGCTCAAGCTGAGAGTGTGAGAGTGTGAGAGCGCTTGTGATGATGATGATGATG 1043
 Db 661 CCGCTCAAGCTGAGAGTGTGAGAGTGTGAGAGCGCTTGTGATGATGATGATGATG 720
 QY 1044 GAGCAGACAC-ACCATGAGATCTGATGATGATGATGATGATGATGATGATGATGATG 1101
 Db 721 GAGCAGACACAAAGGAGATCTGATGATGATGATGATGATGATGATGATGATGATGATG 780
 QY 1102 AAA-----CGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1156
 Db 781 AAAAAGCTCAAG 840
 QY 1157 CAGAGTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1214
 Db 841 CAGAGTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
 QY 1215 CTTTCTCTGAG-CTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1265
 Db 901 CTTTCTCTGAG 953

RESULT 11 803 bp mRNA linear EST 16-DEC-2002
 LOCUS BY708952 803 bp mRNA linear EST 16-DEC-2002
 DEFINITION BY708952 RIKEN full-length enriched, adult male tongue Mus musculus
 cDNA clone 2310030361 5', mRNA sequence.
 ACCESSION BY708952

KEYWORDS	REFERENCE	TITLE	JOURNAL	COMMENT
BY708952.1 GI:27120147				
EST.				
MUS MUSCULUS (house mouse)				
MUS MUSCULUS				
Euarchyotia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
1 (bases 1 to 803)				
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nkaido, I., Saito, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gotojori, T., Baldarelli, R., Hill, D.P., Bull, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batelov, S., Beisel, K.W., Blake, J.A., Bradd, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Fraser, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedziarski, R.M., King, B.L., Konegaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltas, L., Marchionni, L., McKenzie, L., Mikl, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Petten, G., Pesole, G., Petrovsky, N., Pillai, R., Pontus, J.V., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takekura, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verrardo, R., Wagner, L., Walstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilting, L.G., Wynshaw-Boris, A., Yang, S., Yamanaka, Y., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Okawa, T., Kono, H., Nakamura, M., Sakakume, N., Sato, K., Shiraki, T., Waki, K., Imatani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingana, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.				
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs				
Nature 420, 563-573 (2002)				
12466851				
22354683				
Contact: Yoshihide Hayashizaki				
Laboratory for genome exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute				
The Institute of Physical and Chemical Research (RIKEN)				
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan				
Tel: 81-45-503-9222				
Fax: 81-45-503-9216				
Email: genome-resgsc.riken.go.jp				
URL: http://genome.gsc.riken.go.jp/				
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imatani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawaji, J., Kojima, Y., Kondo, S., Kono, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Nunazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y.				
Direct				
Submission				
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)				
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)				
RIKEN Integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)				
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)				
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.				
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.				

FEATURES	SOURCE	Location/Qualifiers
		1..803
		/organism="Mus musculus"
		/mol_type="mRNA"
		/db_xref="taxon:10090"
		/clone="2310003618"
		/sex="male"
		/tissue_type="tongue"
		/dev_stage="adult"
		/lab_host="SOLR"
		/clone_id="RIKEN full-length enriched, adult male tongue"
		/note="Site_1: XhoI; Site_2: SstI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'
		GAGGAGAGAGATGCCAAGAAGCTTTTTTTTTTTTNN 3'] . cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'
		GAGGAGAGATTCGCGATTAAATTAATATACCCCCCCC 3'] . cDNA was cleaved with XhoI and SstI."
BASE COUNT	172 a 222 c 205 g 204 t	
ORIGIN		
Query Match	18.3%; Score 353.2; DB 14; Length 803;	
Best Local Similarity	99.2%; Pred. No. 2,3e-54;	
Matches	355; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
OY	401 CAGCATGATGCTGTGAAGAAAGGAGAGCTGATGTTCTTGCTCCAGATCAGT 460	
Dd	92 CAGCATATGCTGTGAAGAAAGGAGAGCTGATGTTCTTGCTCCAGATCAGT 151	
OY	461 TGTAGCTCTCTTCATCCATATGTCGGTCCACAGACACTTTCACAGAGGAGAGTCCAG 520	
Dd	152 TGTAGCTCTCTTCATCCATATGTCGGTCCACAGACACTTTCACAGAGGAGAGTCCAG 211	
OY	521 GAGGCCCGTGCGATGCTGTGTCCTTCTGCGGTCAGAGTCTCTTTGTGGACA 580	
Dd	212 GAGGCCCGTGCGATGCTGTGTCCTTCTGCGGTCAGAGTCTCTTTGTGGACA 271	
OY	581 GCTTTTGGGAGAGACC GGATGTGTTCTACCTATGAGAGCTGCCTGGGATGTGTGGAT 640	
Dd	272 GCTTTTGGGAGAGACC GGATGTGTTCTACCTATGAGAGCTGCCTGGGATGTGTGGAT 331	
OY	641 GACTTTCACGACGACAGCCGTGGAAGCTGACATGGCTGTGGGATCTTTCGCTTC 700	
Dd	332 GACTTTCACGACGACAGCCGTGGAAGCTGCCATGGCTGTGGGATCTTTCGCTTC 391	
OY	701 CGTCTTCTGCTGTACATGAGCGTCTTTGATGCTCAATGAACCCAGGCCCGGGAAA 758	
Dd	392 CGTCTTCTGCTGTACATGAGCGTCTTTGATGCTCAATGAACCCAGGCCCGGGAAA 449	
RESULT 12		
LOCUS	AL709927	
DEFINITION	DKFZPB8602364_r1 666 (synonym: hlc3) Homo sapiens cDNA clone	EST 12-JUN-2003
ACCESSION	AL709927	
VERSION	AL709927.1 GI:19693282	
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Eukaryota; Metacoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	Bahr,A., Lauber,J., Mewes,H.W., Well,B., Amid,C., Osanger,A., Fobo ,G., Han,M. and Wiemann,S.	
AUTHORS	Est (Bahr,A., Lauber,J., Mewes,H.W., Well,B., et al.)	
TITLE		

JOURNAL
COMMENT

Unpublished
Contact: Bahr A

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
MIPS
This is the 5' sequence of the clone insert
clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Olagen (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.

No s1 sequence available.
This clone (DKFZp68602364) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
source

Location/Qualifiers

1.668
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp68602364"
/issue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/clone_id="686 (synonym: hlcc3)"
/note="Vector: pTRIPLEX2; Site_1: sf1A; Site_2: sf1B;
cDNA-collection"

BASE COUNT 132 a 202 c 178 g 154 t 2 others
ORIGIN

Query Match 18.0%; Score 346; DB 9; Length 668;
Best Local Similarity 76.2%; Pred. No. 4.9e-53;
Matches 451; Conservative 0; Mismatches 137; Indels 4; Gaps 2;

388 TCTCCCTCTCCGAGATGCTGTGTAAGAAAGGAGGCTGCTATGCTGCTG 447
77 TCTTCCACTTCAGACATGCTACTGCTTAAATAAGAGCTCTGCTGTGCTT 136
448 CCCAGGTCATCTGTAGCTCTCTCATTCATATGTCGTCACAGACACTTCCAG- 506
137 CCCAGATGGCCATCTTGCTCTATCTTCCACATGTACACCAACATCAGCTCCCTGT 196
507 --AGGAGAGAGTCCAGAGAGCCGCTGATGCTGCTGCTCTTCTGCGGCTGAGAT 564
197 CTATGAAGAGGACAGCCGAGCGATGACGCTGCTGCTCTCTGCGCTGCTGCT 256
565 CCTCTTTTGGGAGACACTTTTGGGAGACACCGGATGCTTCTACATGAGAGCTG 624
257 CTCTTTTGGGAGACACTTTTGGGAGACACCGGATGCTTCTACATGAGAGCTG 316
625 CCTGGCATGTGTGATGATCTTTCACAGACAGCAGCTGGAAGCTGACATGGCTGTC 684
317 CTTGACAGCTGTGATGATGATCTTTCACAGACAGCAGCTGGAAGCTGACATGGCTGTC 376
685 GGGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 744
377 GGGATCTGATGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 436
745 CAGGCCCCCGAAGACAGTCCAGCCTCTTCCAGTGGGAGCAAGAGCGGCGCTGCTGCTG 804
437 CTGTGCTCCCGAGACAGTCCAGCCTCTTTCAGTGGGAGCAAGAGCGGCGCTGCTGCTG 496
805 CGCCTGTGTGATCTTCTTCCAGTGGGAGCAAGAGCTGCACTGAC-CCAAGCACTGCAAGCTG 863
497 CACCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 556
864 CTCTGGGCTGAGACAGCCCTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 923
557 CTGTGCTGATGAGAGCCCTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 616
924 GTAATCAAGAGAGTGTGCTTTCAGAGCTGAGAGCGCTATGCACTGATCA 975
617 GTGCTCAAGAGAGTGTGCTTTCAGAGCTGAGAGCGCTATGCACTGATCA 668

RESULT 13
ID HS082755 standard; RNA; EST; 583 BP.

AC BX490456;

SV BX490456.1

DT 09-MAY-2003 (Rel. 75, Created)

DE 09-MAY-2003 (Rel. 75, Last updated, Version 1)

XX Homo sapiens mRNA; EST DKFZp6860680_r1 (from clone DKFZp6860680)

XX EST; expressed sequence tag.

XX Homo sapiens (human)

XX Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;

XX Eutheria; Primates; Catarrhini; Hominidae; Homo.

XX Submitted (07-MAY-2003) to the EMBL/GenBank/DBJ databases.

XX MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY

XX This is the 5' sequence of the clone insert

XX Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

XX Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

XX sequenced by EMBL (European Molecular Biology Laboratories,

XX Heidelberg/Germany) within the cDNA sequencing consortium of

XX the German Genome Project.

XX No s1 sequence available.

XX This clone (DKFZp6860680) is available at the RZPD in Berlin.

XX Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,

XX 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

XX Key Location/Qualifiers

XX 1.583

XX /db_xref="taxon:9606"

XX /mol_type="mRNA"

XX /organism="Homo sapiens"

XX /clone="DKFZp6860680"

XX /clone_id="686 (synonym: hlcc3). Vector pSport1_Sf1; host

XX DH10B; sites sf1A + sf1B

XX /dev_stage="adult"

XX /issue_type="cDNA-collection"

SO sequence 583 BP; 120 A; 172 C; 154 G; 137 T; 0 other;

Query Match 15.8%; Score 305; DB 2; Length 583;

Best Local Similarity 76.0%; Pred. No. 1.3e-45;

Matches 390; Conservative 0; Mismatches 120; Indels 3; Gaps 1;

388 TCTCCCTCTCCGAGATGCTGTGTAAGAAAGGAGGCTGCTATGCTGCTG 447
71 TCTTCCACTTCAGACATGCTACTGCTTAAATAAGAGCTCTGCTGTGCTT 130
448 CCCAGGTCATCTGTAGCTCTCTCATTCATATGTCGTCACAGACACTTCCAG- 506
131 CCCAGATGGCCATCTTGCTCTATCTTCCACATGTACAGCAGCAACATCAGCTCCCTGT 190
507 --AGGAGAGAGTCCAGAGAGCCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 564
191 CTATGAAGAGGACAGCCGAGCGCATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 250
565 CCTCTTTTGGGAGACACTTTTGGGAGAGACCGGATGCTTCTACATGAGAGCTG 624
251 CTCTTTTGGGAGACACTTTTGGGAGAGACCGGATGCTTCTACATGAGAGCTG 310
625 CTTGACAGCTGTGATGATGATCTTTCACAGACAGCAGCTGGAAGCTGACATGGCTGTC 684

THIS PAGE BLANK (USPTO)

[illegible]

```

RESULT 3
US-09-927-602-1
: Sequence 1, Application US/09927602
: Patent No. US20020061562A1
: GENERAL INFORMATION:
: APPLICANT: Fukuda, Michiko N.
: APPLICANT: Akama, Tomoya O.
: TITLE OF INVENTION: Methods of Treating Macular Corneal
: TITLE OF INVENTION: Dystrophy
: FILE REFERENCE: P-LI 4852
: CURRENT APPLICATION NUMBER: US/09/927, 602
: CURRENT FILING DATE: 2001-08-09
: PRIOR APPLICATION NUMBER: US 09/638, 211
: PRIOR FILING DATE: 2000-08-11
: NUMBER OF SEQ ID NOS: 38
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 2544
: TYPE: DNA
: ORGANISM: Homo Sapien
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (693)...(1877)
US-09-927-602-1

```

	Query Match	Similarity	19.2%	Score	370.6	DB	9	length	2544
	Best Local	Similarity	63.2%	Pred	No.	3e-108			
	Matches	641	Conservative	0	Mismatches	359	Indels	15	Gaps
QY	524	GCCCGTCATGTCGTGTCGTCTTCCTCGGCGATCAGAGTCCTTTTGTGGGACAGCT	583						
Db	809	GCAGTCATGTGCTGTGTCGTCTCTCGGAGGCGCTCGGAGCTGCTCTTCGTGGGCACT	868						
QY	584	TTTGGGAGACACCCGGATGTGTTTACTATATGAGAGCTCGCTCGCATGTGGATGAC	643						

Db	CTTCAACAGACACCCGACGCTCTTCACTCAATGAGAGCCGCGTGGACAGCTGTGGACAC	928
Db	869	
OY	TTTCAACGACGACACAGCCTGGAAAGCTGCACATGCTGTGGGATCTTCTGCTCCGT	703
Db	929	
OY	CTTCCCTGTGACAGTGAAGCGCTTGTGATGAGCTTACATGAACCCAGAGCCCGGAAACAGTC	763
Db	989	
OY	764	
Db	1046	
OY	824	
Db	1106	
OY	884	
Db	1166	
OY	944	
Db	1226	
OY	1004	
Db	1286	
OY	1064	
Db	1346	
OY	1124	
Db	1403	
OY	1184	
Db	1463	
OY	1241	
Db	1523	
OY	1301	
Db	1583	
OY	1355	
Db	1643	
OY	1415	
Db	1703	
OY	1475	
Db	1763	

RESULT 4
US-09-927-602-38
; Sequence 38, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852


```

; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadamatsu, Kenji
; APPLICANT: Kadamatsu, Kenji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-
; TITLE OF INVENTION: SULFOTRANSFERASE AND
; TITLE OF INVENTION: DNA ENCODING THE SAME
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/10/212.933
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US/09/263,023
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: JP 10-54007
; PRIOR FILING DATE: 1998-03-05
; PRIOR APPLICATION NUMBER: JP 10-177844
; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 2409
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (390)...(1841)
; US-10-212-933-3

```

```

Query Match          4.9%; Score 94; DB 14; Length 2409;
Best Local Similarity 50.7%; Pred. No. 6.4e-19;
Matches 288; Conservative 0; Mismatches 265; Indels 15; Gaps 2;

```

```

QY 535 TCGTGGCTGCTCTTCTGCGGCTCAGAGCTCTTTTGGGACAGCTTTGGGCGAGC 594
DB 751 TGTACGTGTTCACACACGTGGCGCTCTGCTGCTTCTTCTGCGAGCTATTTCAACGAGA 810
QY 595 ACCGGAGTGTCTTACTGTATGAGCGCTGCGCATGTGTGATGACTTTACACGACA 654
DB 811 ATCCGAGGTGTCTTCTCTACGAGCAGTGTGCAATGTATGCAAAAACCTGTATCCGG 870
QY 655 GCACAGCCTGGAACCTGCATGTGCTGCGGATCTTTCGCTTCCGTTCTTCCCTGTG 714
DB 871 GGGAGCGCGTTTCCCTGAGGGGCGAGGCGGAGCATGTGAGCGCTTTTACCGGTGCG 930
QY 715 ACATGAGCGTCTTGTATGCGCTACATGAACCCAGGCCCGG-----AAACAGT 762
DB 931 ACCCTCTGTCTTCCAGTGTGTATAGCCCGCGGCGAGCGGGGCGCAACCTCACACGC 990
QY 763 CCAGCCTCTTCCAGTGGGAGCAAGCGGCGCTGTGCTCAGCGCTGTGTGACTTCT 822
DB 991 TGGGCACTTTTGGGCGACCCACCAAGGTGTGTGTCTGCTACCACTTGTCCCGCT 1050
QY 823 TCCCTGCCACGAGATCAGCTCACCCAAAGCACTCAAGCT---GCTGTGCGGTACGAGC 879
DB 1051 ACCGCAAGAGGAGTGTGGGGTGTGTGAGCAGACCGCGTGTGCAAAAGTGGCCGACAGC 1110
QY 880 CTTTGTATGTGTGAGAGAGCGCTGCGCTTCAAGCGTTGTGTGTACTCAAGAGGTGC 939
DB 1111 GCCGTGGCGCTTTTGGAGAGAGTGCAGTACCGCACACTAGTCAATAAGGCGT 1170
QY 940 GTTTTCTCAGGCTCAGGCGCTTATCCACTACACAGGACCTTCCCTCAACCTGACG 999
DB 1171 GCGCTTTCGACGTGGCGGTCTTGGCGCACTGTGCGAGACCCGCGCTGAGACTCAAGG 1230
QY 1000 TCGTGACACTGTGTCGAGACCCCGGCGCGTGTTCGATCCCGGAGACACCACTAAG 1059
DB 1231 TCATCCACTGTGTGTCGATGCCGCGGCGGTGGGAGTTCACGAGTCCGCTGGGCGACG 1290
QY 1060 AACTCATGTGTGACAGTCAATTTGTGCT 1087
DB 1291 GCTCATTCCTGGAGAGCTTACAGGTGT 1318

```

RESULT 7
US-10-212-933-1

; Sequence 1, Application US/10212933
; Publication No. US2003008366A1

; GENERAL INFORMATION:

; APPLICANT: Uchimura, Kenji

; APPLICANT: Muramatsu, Hideki

; APPLICANT: Kadamatsu, Kenji

; APPLICANT: Kadamatsu, Kenji

; APPLICANT: Habuchi, Osami

; APPLICANT: Muramatsu, Takashi

; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-

; TITLE OF INVENTION: SULFOTRANSFERASE AND

; TITLE OF INVENTION: DNA ENCODING THE SAME

; FILE REFERENCE: TOYAMA1.001AUS

; CURRENT APPLICATION NUMBER: US/10/212.933

; PRIOR FILING DATE: 2002-08-05

; PRIOR APPLICATION NUMBER: US/09/263,023

; PRIOR FILING DATE: 1999-03-05

; PRIOR APPLICATION NUMBER: JP 10-54007

; PRIOR FILING DATE: 1998-03-05

; PRIOR APPLICATION NUMBER: JP 10-177844

; PRIOR FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 2150

; TYPE: DNA

; ORGANISM: Mus musculus

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (470)...(1918)

US-10-212-933-1

```

Query Match          4.6%; Score 88.6; DB 14; Length 2150;
Best Local Similarity 50.1%; Pred. No. 3.3e-17;
Matches 285; Conservative 0; Mismatches 269; Indels 15; Gaps 2;

```

```

QY 534 GTGCTGTGCTGCTTCTTCTGCGGCTCAGAGCTCTTTTGGGACAGCTTTGGGCGAG 593
DB 827 GTGTATGTGTTCACACACGTGGCGCTGCTGCTTCTTCTTCTGAGCTTCTTCAACGAG 886
QY 594 CACCGGAGTGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 653
DB 887 AACCTGAGGTGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 946
QY 654 AGCAGAGCTGTGAGCTGCATGTGTGCGGAGTCTTCTTCTTCTTCTTCTTCTTCTTCT 713
DB 947 GGGAGCGCGTTTCCCTGAGGGGCGAGCGGCGGAGCATGTGAGCGCTTCTTACCGGTGC 1006
QY 714 GACATGAGCTTCTTGTATGCTTACATGTAACCCAGGCCCGCGGAAACAGTCC----- 764
DB 1007 GATCTTTCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1066
QY 765 ---AGCCTCTTCCAGTGGGAGCAAAAGCGGCGCTGTGTGAGGCGCTGTGTGACTTC 821
DB 1067 CTGGGCACTTGTGGGCGACCCACTTACAGAGTGTGTGTCTTCTTCTTCTTCTTCTTCT 1126
QY 822 TTTCCCTGCCACGAGATCAGCTCACCCAAAGCACTGCAAGTGTCTC---TGGGTCAGAGC 878
DB 1127 TACGCGAAGAGAGTGTGTGAGTGTGTGAGAGACCGCGTGTGCAAAAAGTGTGCAACCTCA 1186
QY 879 CCGTTTGTATGTGTGAGAGAGCGCTGCGCTCTCACAGCGCTTCTTCTTCTTCTTCTTCT 938
DB 1187 CGCGTGGACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1246
QY 939 CGTTTCTCAGCTGCAAGAGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 998
DB 1247 CGGGCTTCTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1306
QY 999 GTCTGTGACCTGTGTGTGAGAGCCCGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1058

```

Db 1307 GTCATCACCCTAGCTAGCTGTCCTGCTGCTGCGACGCTCCGCAATCCGCTGCTCAC 1366
 QY 1059 GAACATGATGTTGACAGTATATTTGCT 1087
 Db 1367 GGCCTCATCCGGGAAAGCCTACAGGTGCT 1395

RESULT 8

US-09-735-705-63
 ; Sequence 63, Application US/09735705
 ; Patent No. US20020052329A1
 ; GENERAL INFORMATION:

APPLICANT: Mang, Tongtong
 APPLICANT: Pan, Liqun
 APPLICANT: Kalos, Michael D.
 APPLICANT: Bangur, Chaitanya S.
 APPLICANT: Hosken, Nancy
 APPLICANT: Fanger, Gary R.
 APPLICANT: Li, Samuel X.
 APPLICANT: Wang, Aljun
 APPLICANT: Skelky, Yasir A.W.
 APPLICANT: Henderson, Robert A.
 APPLICANT: McNeill, Patricia D.
 APPLICANT: Fanger, Neil
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
 FILE REFERENCE: 210121.455C14
 CURRENT APPLICATION NUMBER: US/09/735,705
 CURRENT FILING DATE: 2000-12-12
 NUMBER OF SEQ ID NOS: 419
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 63
 LENGTH: 731
 TYPE: DNA
 ORGANISM: Homo sapien
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (1)...(731)
 OTHER INFORMATION: n = A,T,C or G
 US-09-735-705-63

Query Match 2.7%; Score 52.4; DB 9; Length 731;
 Best Local Similarity 57.2%; Pred. No. 8.2e-06;
 Matches 95; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 922 TGGTACTCAAGAGGTGCGTTTCTCAGCTGCGAGCCCTATCCACTACTACGAGCC 981
 Db 3 TAGTATAAAGGTTGCGCGCTTCGAGCTGGCGCTTTGGCGCCACTGCTCGAGACC 62
 QY 982 CTTCCTCAACCTGCAGCTGCTGACCTGCTCCGAGACCCCGGGCGCTTCCGATCCC 1041
 Db 63 CGGCCCTGGACCTCAAGCTCATCCACTTGTGCTGATCCCGCGCGGAGGATTCAC 122
 QY 1042 GGGAGACACCAACATAGAACATGATGTCAGACATATTTGCT 1087
 Db 123 GGATCCGCTCGCGCCAGCGCTCATCCGAGAGAGCTACAGGTGCT 168

RESULT 9

US-09-850-716A-63
 ; Sequence 63, Application US/09850716A
 ; Patent No. US20020115139A1
 ; GENERAL INFORMATION:

APPLICANT: Kalos, Michael D.
 APPLICANT: McNeill, Patricia D.
 APPLICANT: Retter, Marc W.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
 FILE REFERENCE: 210121.455C15
 CURRENT APPLICATION NUMBER: US/09/850,716A
 CURRENT FILING DATE: 2001-05-07
 NUMBER OF SEQ ID NOS: 440

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 63
 LENGTH: 731
 TYPE: DNA
 ORGANISM: Homo sapien
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (1)...(731)
 OTHER INFORMATION: n = A,T,C or G
 US-09-850-716A-63

Query Match 2.7%; Score 52.4; DB 10; Length 731;
 Best Local Similarity 57.2%; Pred. No. 8.2e-06;
 Matches 95; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 922 TGGTACTCAAGAGGTGCGTTTCTCAGCTGCGAGCCCTATCCACTACTACGAGCC 981
 Db 3 TAGTATAAAGGTTGCGCGCTTCGAGCTGGCGCTTTGGCGCCACTGCTCGAGACC 62
 QY 982 CTTCCTCAACCTGCAGCTGCTGACCTGCTCCGAGACCCCGGGCGCTTCCGATCCC 1041
 Db 63 CGGCCCTGGACCTCAAGCTCATCCACTTGTGCTGATCCCGCGCGGAGGATTCAC 122
 QY 1042 GGGAGACACCAACATAGAACATGATGTCAGACATATTTGCT 1087
 Db 123 GGATCCGCTCGCGCCAGCGCTCATCCGAGAGAGCTACAGGTGCT 168

RESULT 10

US-09-897-778-63
 ; Sequence 63, Application US/09897778
 ; Patent No. US20020147143A1
 ; GENERAL INFORMATION:

APPLICANT: Wang, Tongtong
 APPLICANT: Marnerakis, Margarita
 APPLICANT: Fanger, Gary R.
 APPLICANT: Vedvick, Thomas S.
 APPLICANT: Carter, Darrick
 APPLICANT: Watanabe, Yoshihiro
 APPLICANT: Henderson, Robert A.
 APPLICANT: Peckham, David W.
 APPLICANT: Fanger, Neil
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
 FILE REFERENCE: 210121.455C16
 CURRENT APPLICATION NUMBER: US/09/897,778
 CURRENT FILING DATE: 2001-06-28
 NUMBER OF SEQ ID NOS: 467
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 63
 LENGTH: 731
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 237, 249, 263, 288, 312, 317, 323, 326, 337, 352, 362, 370,
 LOCATION: 377, 400, 411, 414, 434, 436, 446, 457, 473, 486, 497, 498,
 LOCATION: 502, 512, 531, 546, 554, 563, 565, 566, 588, 608, 611,
 LOCATION: 613, 615, 627, 632, 640, 641, 644, 654, 660, 663, 665
 OTHER INFORMATION: n = A,T,C or G
 NAME/KEY: misc_feature
 LOCATION: 671, 678, 692, 697, 698, 699, 704, 705, 712, 714, 717, 718,
 LOCATION: 719, 723, 725, 730, 731
 OTHER INFORMATION: n = A,T,C or G
 US-09-897-778-63

Query Match 2.7%; Score 52.4; DB 10; Length 731;
 Best Local Similarity 57.2%; Pred. No. 8.2e-06;
 Matches 95; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 922 TGGTACTCAAGAGGTGCGTTTCTCAGCTGCGAGCCCTATCCACTACTACGAGCC 981
 Db 3 TAGTATAAAGGTTGCGCGCTTCGAGCTGGCGCTTTGGCGCCACTGCTCGAGACC 62

DB 123 GCTCGCGCCAGCGCTCATCCGTGAGAGCCTACAGTGTGT 162

RESULT 14

US-09-833-790-401

; Sequence 401, Application US/09833790
; Patent No. US20020068288A1

GENERAL INFORMATION:

; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tonglong
; APPLICANT: Secrist, Heather
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Indritas, Carol Y.

; APPLICANT: Fan, Liqun

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.512

; CURRENT APPLICATION NUMBER: US/09/833,790

; CURRENT FILING DATE: 2001-04-11

; NUMBER OF SEQ ID NOS: 440

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 401

; LENGTH: 303

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-833-790-401

Query Match 2.7%; Score 51.2; DB 9; Length 303;
Best Local Similarity 57.5%; Pred. No. 1.1e-05;

Matches 92; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

DB 928 TCAAGAGGTCGCTTCTCAAGCCTGCAGCCCTCTATCCACTACTACGAGACCCCTTCCC 987

DB 3 TAAAGGCTGTGCGCTCTTGCAGCGTGGCGCTTGGCCCACTGCTGCGAGACCCGCGCC 62

OY 988 TCAACCTGACGTGCGACCGTGGTCCGAGACCCCGGCGCTTCCATCCCGGAGC 1047

DB 63 TGGACCTCAAGGTCATCCACTTGTGCGTGATCCCGCGGCTGGCGAGTTCACGAGATCC 122

OY 1048 ACACCAACCATAGACTGATGTTGACAGTCAATATGTGCT 1087

DB 123 GCTCGCGCCAGCGCTCATCCGTGAGAGCCTACAGTGTGT 162

RESULT 15

US-10-240-485-175

; Sequence 175, Application US/10240485
; Publication No. US20030148327A1

GENERAL INFORMATION:

; APPLICANT: OLEK, Alexander

; APPLICANT: PIEPERROCK, Christian

; APPLICANT: BERLIN, Kurt

; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: Metastasis

; FILE REFERENCE: 5013.1007

; CURRENT APPLICATION NUMBER: US/10/240,485

; CURRENT FILING DATE: 2002-10-02

; PRIOR APPLICATION NUMBER: PCT/EP01/03970

; PRIOR FILING DATE: 2001-04-06

; PRIOR APPLICATION NUMBER: DE 10019058.8

; PRIOR FILING DATE: 2000-04-06

; PRIOR APPLICATION NUMBER: DE 10019173.8

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: DE 10032529.7

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: DE 10043826.1

; PRIOR FILING DATE: 2000-09-01

; NUMBER OF SEQ ID NOS: 202

; SEQ ID NO 175

; LENGTH: 23683

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-175

Query Match 2.2%; Score 43.2; DB 12; Length 23683;

Best Local Similarity 51.6%; Pred. No. 0.07;
Matches 99; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

OY 105 GTGATTAATGTGCTGTAGATATTCGCTGCTGTGATGATGAGAAAGCCAGAGG 164

DB 13449 GGGTTTATGTGTGTGGTAGAATAGTTTTCATTTTCGATTTGTTATTTT 13508

OY 165 GTGAGTTTAAAGACTTGTACATAGATGACGTGATCCAAATTAAGCCAGAAATTA 224

DB 13509 GGTTCGAGAGTGTGGATTAATAGCGCTGATGATCGATTCGATTAATGTTTAA 13568

OY 225 TGCAGGAGATGTGCAAAATACCTCCAGAGATTTTGTGTTTGTGTTGTTTGT 284

DB 13569 TTTTTRAGAAAAATAAAATTAATTAAGAAATTAATTTTATTTTATTTGATTTGTTG 13628

OY 285 TTGTTTACATT 296

DB 13629 GTGTTATTAATT 13640

Search completed: August 14, 2003, 18:01:02
Job time : 471.88 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 11:31:05 ; Search time 124.227 Seconds

(Without alignments)
6843.163 Million cell updates/sec

Title: US-09-645-078-3

Perfect score: 1926
Sequence: 1 ggcgcactactactactg.....ccgtcactcttcttaag 1926

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 56978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NH:*
1: /cgn2_6/pdata1/1/ina/5A_COMB.seq:*
2: /cgn2_6/pdata1/1/ina/5B_COMB.seq:*
3: /cgn2_6/pdata1/1/ina/6A_COMB.seq:*
4: /cgn2_6/pdata1/1/ina/6B_COMB.seq:*
5: /cgn2_6/pdata1/1/ina/PCRTUS_COMB.seq:*
6: /cgn2_6/pdata1/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	741.6	38.5	2032	US-09-045-284A-1	Sequence 1, Appl1
2	741.6	38.5	2032	US-09-190-911-2	Sequence 2, Appl1
3	717	37.2	2065	US-09-786-240-26	Sequence 26, Appl1
4	114	5.9	2190	US-09-015-188-1	Sequence 1, Appl1
5	107.4	5.6	2156	US-08-899-514-1	Sequence 1, Appl1
6	98.4	5.1	2354	US-08-655-878-1	Sequence 1, Appl1
7	94	4.9	2409	US-09-263-023-3	Sequence 3, Appl1
8	94	4.9	2409	US-09-471-867-3	Sequence 3, Appl1
9	88.6	4.6	2150	US-09-263-023-1	Sequence 1, Appl1
10	88.6	4.6	2150	US-09-471-867-1	Sequence 1, Appl1
11	52.4	2.7	731	US-09-040-984-63	Sequence 63, Appl1
12	52.4	2.7	731	US-09-123-912-63	Sequence 63, Appl1
13	52.4	2.7	731	US-09-643-597-63	Sequence 63, Appl1
14	52.4	2.7	731	US-09-480-884A-63	Sequence 63, Appl1
15	52.4	2.7	731	US-09-542-615A-63	Sequence 63, Appl1
16	52.4	2.7	731	US-09-606-421B-63	Sequence 63, Appl1
17	45.2	2.3	7218	US-08-232-463-14	Sequence 14, Appl1
18	40.4	2.1	4352	US-09-620-312D-383	Sequence 383, Appl1
19	40	2.1	1233	US-09-266-965-24	Sequence 24, Appl1
20	40	2.1	1233	US-09-266-965-24	Sequence 24, Appl1
21	40	2.1	18331	US-09-266-965-96	Sequence 96, Appl1
22	39.8	2.1	1269	US-08-808-641-2	Sequence 2, Appl1
23	39.8	2.1	1269	US-09-064-839-2	Sequence 2, Appl1
24	39.8	2.1	1269	US-09-351-438-2	Sequence 2, Appl1
25	37.8	2.0	12597	US-09-705-299-12	Sequence 12, Appl1
26	37.6	2.0	1599	US-08-853-733B-1	Sequence 1, Appl1
27	37.6	2.0	1603	US-08-625-209A-1	Sequence 1, Appl1

28	37.6	2.0	1603	US-08-675-885-6	Sequence 6, Appl1
29	37.6	2.0	1847	US-08-675-885-4	Sequence 4, Appl1
30	37	1.9	1488	5268463-6	Patent No. 5268463
31	36.6	1.9	289	US-09-007-005-17	Sequence 17, Appl1
32	36.6	1.9	289	US-09-244-796-17	Sequence 17, Appl1
33	36.6	1.9	289	US-08-894-997-49	Sequence 49, Appl1
34	36.4	1.9	1282	US-08-878-989-12	Sequence 12, Appl1
35	36.4	1.9	1282	US-09-272-796-12	Sequence 12, Appl1
36	36.4	1.9	1282	US-09-016-434-953	Sequence 953, Appl1
37	36.4	1.9	3223	US-07-980-528-1	Sequence 1, Appl1
38	35.4	1.8	990	US-09-252-991A-11517	Sequence 11517, A
39	35.4	1.8	2004	US-09-252-991A-11865	Sequence 11865, A
40	35.4	1.8	2166	US-09-252-991A-11722	Sequence 11722, A
41	35.2	1.8	1506	US-09-176-657-5	Sequence 5, Appl1
42	35.2	1.8	1506	US-09-421-299-5	Sequence 5, Appl1
43	35.2	1.8	1872	US-09-252-991A-9239	Sequence 9239, Ap
44	35	1.8	1287	US-09-252-991A-2964	Sequence 2964, Ap
45	35	1.8	1457	US-09-620-312D-408	Sequence 408, App

ALIGNMENTS

RESULT 1	US-09-045-284A-1	Sequence 1, Application US/09045284A
Patent No. 6265192		
GENERAL INFORMATION:		
APPLICANT:	Bistrup, Annette	
APPLICANT:	Rosen, Steven D.	
APPLICANT:	Hammerlich, Stefan	
TITLE OF INVENTION:	GLYCOSYL SULFOTRANSFERASE-3	
FILE REFERENCE:	6510-107051	
CURRENT APPLICATION NUMBER:	US/09/045,284A	
CURRENT FILING DATE:	1998-03-20	
NUMBER OF SEQ ID NOS:	9	
SOFTWARE:	FastSeq for Windows Version 4.0	
SEQ ID NO 1		
LENGTH:	2032	
TYPE:	DNA	
ORGANISM:	Homo sapiens	
US-09-045-284A-1		
Query Match	38.5%	Score 741.6; DB 3; Length 2032;
Best Local Similarity	76.4%	Pred. No. 2.7e-220;
Matches 937; Conservative	0; Mismatches 284; Indels	5; Gaps 2;
QY	388	TCCTCTTCCGAGGATGCTGTGAAGGAGAGGCTGATGCTCTGGGTT 447
DB	165	TCCTCCACTTCAGACAAATGCTACTGCTAAAGGCTCTCTGTTGTTT 224
QY	448	CCGAGTCATGCTGTGCTCTTCATCATATGCTCCGACAGACACCTTCCAG- 506
DB	225	CCGAGATGAGGATCTGCTCTATCTTCACATGTACACCAACATAGCTCCCTGT 284
QY	507	--AGGAGAGAGTCAGAGAGCCGCGATGCTGCTGCTCTCTGCGGTCAGAT 564
DB	285	CTATGAAGGACACACCCGACGATGCTGCTGCTCTCTGCGGTCAGAT 344
QY	565	CCCTCTTGGGACACCTTTCGGGACAGACCCGATGCTTCTTACCTGATGAGCTG 624
DB	345	CTTCTTGGGACACCTTTCGGGACAGACCCGATGCTTCTTCTGATGAGCTG 404
QY	625	CTGCGATGCTGATGCTTTCACAGACAGACAGCTGAGACCTGACATGCTGCTGC 684
DB	405	CTGCGACGCTGATGCTTTCACAGACAGACAGCTGATGCTGATGCTGCTGC 464
QY	685	GGATCTTGGCTGCT 744
DB	465	GGATCTGATGAGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 524
QY	745	CAGCCCGGAGAACTGACCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 804

Db	525	CTGGTCCCGGAGACAGTCAGCCCTCTTTCAGTGGAGAACAGCGCGGCCCTCGTGTCTG	584
QY	805	CGCCTGTGTGACTTCTTCCCTGCCACAGATATAGTCAACCAAGCATCTGCAGCTGC	864
Db	585	CACCTGCTGTGACATCATCTCCACAAAGATGAATCATCCCCGGGCTCACTGCAGGCTCC	644
QY	865	TCTCGGCTACACAGCCCTTTTATATATGTGTGAGAAAGGCGCTGCAGGCTCTCAGGCTCTG	924
Db	645	TGTGACATCAACAGCGCTTTGAGGTTGGTGGAGAAAGCGCTGCAGCTCTCAAGCCAGCTGC	704
QY	925	TACTCAAGGAGTGGCTTTTCTCAGACCTGCAGGCGCTTATTCACACTACTCAGGACCTT	984
Db	705	TGCTCAAGGAGAGTGCCTTCTTTCACAAACCTGCAGCTCCCTTACCCGCTGTGAAGACCCCT	764
QY	985	CCCTCAACCTCAGCTGCTGTCACCTGTGTCAGAGACCCCGGCGCTGTTCGATCCGCGG	1044
Db	765	CCCTCAACCTCAGATATCTGTCACCTGTGTCAGAGACCCCGGCGCTGTTCGATCCGCGG	824
QY	1045	AGCACACCACCATAGAACTCATGGTTGACAGTCATATTTGCTAGAGGAGCATTTTGA	1104
Db	825	AACCCACAAAGGAGATCTCATGTATGTACATCGCATGTATGTATGGGCGACATGAGCAGA	884
QY	1105	CGATCAAGAGAGAAAGACAGCCCATATTTGTCATGAAATCATGTCGCAAAAGCCAGAGTG	1164
Db	885	AATCTCAAGAGAGAGACCAACCCCTACATCTGTGTGATGTCAGGTATCTGCCAAAGCCAGCTG	944
QY	1165	ACATAGTCAGAGCCATCCAAACCCCTCCCTGAAGCTCTGCAGACAGCGCTACCTGTTCTGA	1224
Db	945	AGATCTCAACAAACCATCAGCTCCTTGTGCCAAGGCCCTCAGAAACGCTACCTGCTTGTGC	1004
QY	1225	GGTATGAGGAGCTGTGGTGGGCGCCCGGCGCCAGACAGACAGCATATATATATTTGTG	1284
Db	1005	GCTATGAGAGACTGTGCTGAGCCCTGTGGCCACAGCTTCCCGAATGTATGAATTCGTGG	1064
QY	1285	GGTTGGATTTTTCGCCACCTCCAAACATGGGTTTACATATGTCAACCCGCGCAGAGGCA	1344
Db	1065	GATTGGAATTTTCTCCCATCTTTCAGACTGGTGGTGCATATACATCAACCCGAGCAAGGCA	1124
QY	1345	TGGGTCAAGCATCCCTTCATACTAAGCCAGAGAAAGCCCTCAAGCTCTCTCAGGCGTGGC	1404
Db	1125	TGGGTGACACAGCTTTCACACAAATGCGAGGAGTCCCTTAATGTCTCCAGGCTTGGC	1184
QY	1405	GTTGGTCTTACCTTACGAAAGGTTTCCAGCTTCAAGATGCTCGCGGTGAGGCTATGG	1464
Db	1185	GCCTGTCTTCCCTATGAAAGGTTTCTGACATTCACAAAGCCGTGGCATGCGCATGA	1244
QY	1465	ATTTCGTGAGATACCTTCAGGTCTCAGATCTCAACAAAGCAAGGCAACCTGTCTCTGATC	1524
Db	1245	ATTTCGTGAGGCTACCGCACGTCTGAGATCTGAACAAAGCAAGAAACCTGTTCGTGGATC	1304
QY	1525	TTTCGTCTCTCCATATCTTTGGGGCGAGTCTTCCGAGAGGTTAAGAGGTCTGTGTG	1584
Db	1305	TTTCGTCTCTCCATATCTTCCCTCAGCAAAATTCACATAAGAGGTTGAGCAAGCTTGTCTG	1364
QY	1585	CACCCCTTGGTTCAGACCTTATGTCAC	1610
Db	1365	--CCACCTGGTGCAGCTCAGTCAC	1388
RESULT 2			
US-09-190-911-2			
: Sequence 2, Application US/09190911			
: Patent No. 6365365			
: GENERAL INFORMATION:			
: APPLICANT: Bistrup, Annette			
: APPLICANT: Rosen, Steven D.			
: APPLICANT: Tangemann, Kirsten			
: APPLICANT: Hemmrich, Stefan			
: TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3			
: FILE REFERENCE: 6510-107CIP			
: CURRENT APPLICATION NUMBER: US/09/190,911			
: CURRENT FILING DATE: 1998-11-12			
: EARLIER APPLICATION NUMBER: 09/045,284			

```

; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens
US-09-190-911-2

Query Match      38.5%; Score 741.6; DB 4; Length 2032;
Best Local Similarity 76.4%; Pred. No. 2,7e-224;
Matches 937; Conservative 0; Mismatches 280; Indels 5; Gaps 2.

QY  TCTCCTTTCGCCAGATGATGATGCGTTGGAAGAAAGGAGAGCGTGTGATGTTCCGSGGT 447
Db  TCTTCACCTTCAGCCAAATGCTACTGCTTAAATAAATTAACCTCTCGTGTTCGGTTT 224
QY  448 CCCAGGTCATGTTGTAGCTCTCTTCATCATCATATGTCGTCACAGACACCTTTCACG- 506
Db  225 CCCAGATGGCCATCTTGGCTCCTATTCCTTCCAAATGATACAGCACAAATCATAGCTCCGT 284
QY  507 --AGGAGAGATCCAGAGAGGCCCGGTGCATGTGCTGTCTCTTCTCGGGGGTCAGAT 564
Db  285 CTATGAAGAGCACAGCCCGAGCATGACAGTGCCTGTTCTCTCTCGGGGGCTTGCT 344
QY  565 CCTCTTTTGTAGAGACGTTTTCGGGCGACACCGGATGTGTTCTACCTGATGAGAGCTG 624
Db  345 CTCTCTTTGTGGGAGAGCTTTTGGGCGACACCCAGATGTTTCTACCTGATGAGAGCCG 404
QY  625 CCTGGCATGTGTGATGATCTTTCACACAGCACACAGCCTGGAAGCTGACATGCTGTGC 684
Db  405 CCTGGACAGTGTGATGATGATCTTTCACAGCACAGACACCGCCTGATGCTGCATGCTGTGC 464
QY  685 GGGATCTTCTGCGTTCCGCTTCCGTGTGATGATGAGGCTGTTGATGCCATATGAAAC 744
Db  465 GGGATCTTATACGGGCGCTCTTCTTGTGCGACATGATGAGCTCTTGTATCCTTACATGGAAC 524
QY  745 CAGGCCCCCGGAAACAGTCCAGAGCCTCTTCCAGTGGGAGACAAAGCCGGGCCCTGTGCTAG 804
Db  525 CTGGTCCCCGGAGACAGTCCAGAGCCTCTTCAATGTTGGAATAACAGCCGGGCCCTGTGTTCTG 584
QY  805 CGCCTGTGTGATCTTCTCCCTGCGCCACAGATCAGCTCACACCCAGCAGTGCAGCTGC 864
Db  585 CACCTGCTGTGATCATCATCCACAAAGATGAATATCATCCCCGGGCTCAGTCCAGGCTCC 644
QY  865 TCTGGGTCAGAGAGCCCTTGTATATGATGAGGAAGAGCCCTGCGCTCCTACGGCTCCGTTG 924
Db  645 TGTGAGTCAACAGCCCTTTTGAAGGTGGTGGAGAAAGAGCCCTGCCGCTCTTACAGCCAGCTG 704
QY  925 TACTCAAGAGAGGTGCTTCTTCAGCCCTGACAGGCCCTCTATTCATCTACTACAGGACCTT 984
Db  705 TGCTCAAGAGAGTGGGCTTCTTCAACCTGCAAGTCCCTCTACCCGCTGTGAAAGACCCCT 764
QY  985 CCGTTCACCTGCAGCTGCTGTGACACCTGGTCCGAGAGACCCCGGGCGGTGTCCGATCCCGGG 1044
Db  765 CCGTCAACCTGCATATCTGTGACCTGTGCGGGAGACCCCGGGCGGTGTCCGTTCCCGAG 824
QY  1045 AGCAGACACCATTAAGATCACTAATGTTGACAGTCAATATGTCTAGAGGAGCATTTGGAAA 1104
Db  825 AAGCGACAAAGGAGATCTCATGATGTGACATGTCGATTTGTGATGGGCGACGATGACAGAA 884
QY  1165 ACATATGTAAGGCCATCCAAACCTCTCCTGAAAGCTCTGACAGAGCGCTACTGTTCCTGGA 1224
Db  945 AGATCTCAACAAACCATCCAGTCTTGTGCCAAAGGCCCTGACAGAAAGCTACTGCTTGGTC 1004
QY  1225 GGTAATGAGAGACCTGGTTGGGGACCCCGCGGCGCCAGAGACACAGATATATATAATTTGGG 1284
Db  1005 GCTATGAGAGACCTGGCTGAGCCCTGTGGGCCACAGACTTCCCGAATGTATATAATTCGTGG 1064

```


Db 1500 GGCTACAACTGG 1512

RESULT 6
US-08-655-878-1
; Sequence 1, Application US/08655878

```

GENERAL INFORMATION:
APPLICANT: FUKUTA, MASAKAZU
APPLICANT: HARUCHI, OSAMI
TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSSEE:
STREET:
CITY:
STATE:
COUNTRY:
ZIP:
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/655,878
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE:
TELEFAX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2354
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Chick
TISSUE TYPE: Embryo chondrocyte
FEATURE:
FEATURE: NAME/KEY: CDS
FEATURE: LOCATION: 211..1584
FEATURE: IDENTIFICATION METHOD: P
FEATURE:
FEATURE: NAME/KEY: sig.peptide
FEATURE: LOCATION: 211..309
FEATURE: IDENTIFICATION METHOD: P
FEATURE:
FEATURE: NAME/KEY: mat.peptide
FEATURE: LOCATION: 310..1584
FEATURE: IDENTIFICATION METHOD: P
FEATURE:
FEATURE: NAME/KEY: transmembrane domain
FEATURE: LOCATION: 280..321
FEATURE: IDENTIFICATION METHOD: P
FEATURE:
FEATURE: NAME/KEY: potential N-glycosylation site
FEATURE: LOCATION: 394..402
FEATURE: IDENTIFICATION METHOD: S
FEATURE:
FEATURE: NAME/KEY: potential N-glycosylation site
FEATURE: LOCATION: 427..435
FEATURE: IDENTIFICATION METHOD: S
FEATURE:

```


Db 1051 ACCCAAGAGAGTGTGGGTGGTGGACGACCCGCTGTGCATGAAGAGCCCGCACAGC 1110
QY 880 CCTTGANATGTGGAGAGAGCCCTGCGCTCTACAGGCTTGCTGTACTCAAGAGGTGC 939
Db 1111 GCGTGGGCGCTTGTGGAGAGAGTGGCCAGTACCGACACTAGTCTAAAGGGTGTGC 1170
QY 940 GTTTCATGAGCTCAGAGCCCTTATCCACTACTACAGGACCTTCCCTCAACCTGCAGC 999
Db 1171 GCGTCTTCGAGCTGCGGTCTTGGCGCCACTGTGCGAGACCCGCGCTTGACCTCAAGC 1230
QY 1000 TCGTGACCTGTGTCGAGACCCCGGCGGTTCGATCCGATCCGGGAGACACACACATAG 1059
Db 1231 TCATCCACTGTGTGCTGATCCCGCGGGTGGCGAGTTACAGGATCCGCTCGCGCCACG 1290
QY 1060 AACCATGTGAGACATCATTTGTGCT 1087
Db 1291 GCGTATCCGTGAGAGCCTACAGGTGT 1318

RESULT 8

US-09-471-867-3

Sequence 3, Application US/09471867
Patent No. 6455289

GENERAL INFORMATION:

APPLICANT: Uchimura, Kenji

APPLICANT: Muramatsu, Hideki

APPLICANT: Kadomatsu, Kenji

APPLICANT: Kanagaki, Reiji

APPLICANT: Habuchi, Osami

APPLICANT: Muramatsu, Takashi

TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND

FILE REFERENCE: TOYAMA1.001AUS

CURRENT APPLICATION NUMBER: US/09/471,867

EARLIER FILING DATE: 1999-12-23

EARLIER APPLICATION NUMBER: US 09/263,023

EARLIER FILING DATE: 1999-03-05

EARLIER APPLICATION NUMBER: JP 10-54007

EARLIER FILING DATE: 1998-03-05

EARLIER APPLICATION NUMBER: JP 10-177844

EARLIER FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 10

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 3

LENGTH: 2409

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (390)...(1841)

US-09-471-867-3

Query Match 4.9%: Score 94; DB 4; Length 2409;
Best Local Similarity 50.7%: Pred. No. 1.3e-18;

Matches 288; Conservative 0; Mismatches 265; Indels 15; Gaps 2;

QY 535 TGTGTGCTGTCTTCTTCTGCGGCTCAGATCCTTTTGTGGACAGCTTTTCGGGAGC 594
Db 751 TGTGTGCTGTCTTCTTCTGCGGCTCAGATCCTTTTGTGGACAGCTTTTCGGGAGC 810
QY 595 ACCCGAGTGTCTTCTTCTGCGGCTCAGATCCTTTTGTGGACAGCTTTTCGGGAGC 654
Db 811 ATCCCGAGTGTCTTCTTCTGCGGCTCAGATCCTTTTGTGGACAGCTTTTCGGGAGC 870
QY 655 GACACAGCTGTGAGTGTGCTGTGCGGATCTTCTGCTTCCCTTCTTCTGCTGTG 714
Db 871 GCGAGCGCGTTCCTGTGAGGAGGAGCGGAGCATGTGAGGCTTTTACCGCTGCG 930
QY 715 ACATGAGCGCTTGTGATGCTTACATGAACCCAGGCCCCCGG-----AAACAGT 762
Db 931 ACCTCTGTCTTCTTCTGATGTATAGCCCGCGGCGAGCGGGGCGGCAACCTTACACAGC 990

RESULT 9

US-09-263-023-1

Sequence 1, Application US/09263023
Patent No. 6037159

GENERAL INFORMATION:

APPLICANT: Uchimura, Kenji

APPLICANT: Muramatsu, Hideki

APPLICANT: Kadomatsu, Kenji

APPLICANT: Kanagaki, Reiji

APPLICANT: Habuchi, Osami

APPLICANT: Muramatsu, Takashi

TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND

FILE REFERENCE: TOYAMA1.001AUS

CURRENT APPLICATION NUMBER: US/09/263,023

EARLIER FILING DATE: 1999-03-05

EARLIER APPLICATION NUMBER: JP 10-54007

EARLIER FILING DATE: 1998-03-05

EARLIER APPLICATION NUMBER: JP 10-177844

EARLIER FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 10

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 1

LENGTH: 2150

TYPE: DNA

ORGANISM: Mus musculus

FEATURE:

NAME/KEY: CDS

LOCATION: (470)...(1918)

US-09-263-023-1

Query Match 4.6%: Score 88.6; DB 3; Length 2150;
Best Local Similarity 50.1%: Pred. No. 5.9e-17;

Matches 285; Conservative 0; Mismatches 269; Indels 15; Gaps 2;

QY 534 GTGCGTGTGTCTTCTTCTGCGGCTCAGATCCTTTTGTGGACAGCTTTTCGGGAGC 593
Db 827 GTGTATGTGTTCACACAGCTGCGGCTCAGATCCTTTTGTGGACAGCTTTTCGGGAGC 886
QY 594 CACCGAGTGTCTTCTTCTGCGGCTCAGATCCTTTTGTGGACAGCTTTTCGGGAGC 653
Db 887 AACCTGAGTGTCTTCTTCTGCGGCTCAGATCCTTTTGTGGACAGCTTTTCGGGAGC 946
QY 654 AGCACAGCTGTGAGTGTGCTGTGCGGATCTTCTGCTTCCCTTCTTCTGCTGTG 713
Db 947 GCGGAGCGCGTTCCTGTGAGGAGGAGCGGAGCATGTGAGGCTTTTACCGCTGC 1006
QY 714 GACATGAGCGTCTTGTGATGCTTACATGAACCCAGGCCCCCGGAAACATCC----- 764

Db 1007 GATCTTCGGTTCAGCTGATATACCCCGCAGGAGTGGGGGGGCAACCTCACCAGT 1066
QY 765 ---AGCCTCTTCAGTGGAGCAAAAGCCGGCCCTGTGCTCAGCGCTGTGTACTTC 821
Db 1067 CTGGGATCTTTTGGGAGCAGCAGCTAACAAAGTGTGTGTCTCTGCGCACTCTCTGCC 1126
QY 822 TTCCCTGCCACAGATCAGCTCACCAGCAGCTGCAAGCTGCTC---TGGGTACAGCA 878
Db 1127 TACCGCAAGAGAGTCTCGAGTGTGTGAGACAGCCGGGTGTGCAAAAAGTGGCCACCTCAA 1186
QY 879 CCCTTGATATAGTGTGAGAAAGGCTCCCTCTCAGAGGCTTCTGTGTACTCAAGAGGTG 938
Db 1187 CGCTGGACAGCTTCGAGAGAGGTGTCCAGATACCGCAGCGGTGTATCAAGGGCGTG 1246
QY 939 CGTTTCTGAGCTCGAGGCGCCCTCTATCCACTACTCAGGAGCCCTTCCCTCAACTGCAC 998
Db 1247 CGGGCTTGTGATGTGGCTGTGTGGGCGCCGCTCTTAAGATATCAGCCTTGGACCTCAAG 1306
QY 999 GTGCTGACCTGTGTCCGAGAGCCCGGGCGGCTGTCCGATCCGGAGCACACCCACCA 1058
Db 1307 GTATCCACCTAGTATAGTATCTCTGCTGTGCGAGCTCCCGCATCCGCTCGCTCAC 1366
QY 1059 GAACCTATGTTGACAGTATATTTGCT 1087
Db 1367 GGCCTATCCGGGAAAGCTTACAGGTGT 1395

RESULT 10
US-09-471-867-1
Sequence 1, Application US/09471867
Patent No. 6455289

GENERAL INFORMATION:
APPLICANT: Uchimura, Kenji
APPLICANT: Muramatsu, Hideki
APPLICANT: Kadomatsu, Kenji
APPLICANT: Kannagi, Reiji
APPLICANT: Habuchi, Osami
APPLICANT: Muramatsu, Takashi
TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
FILE REFERENCE: TOYAMA41.001AUS
CURRENT APPLICATION NUMBER: US/09/471,867
EARLIER FILING DATE: 1999-12-23
EARLIER APPLICATION NUMBER: US 09/263,023
EARLIER FILING DATE: 1999-03-05
EARLIER APPLICATION NUMBER: JP 10-54007
EARLIER FILING DATE: 1998-03-05
EARLIER APPLICATION NUMBER: JP 10-177844
EARLIER FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 2150
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (470)...(1918)
US-09-471-867-1

Query Match 4.6%; Score 88.6; DB 4; Length 2150;
Best Local Similarity 50.1%; Pred. No. 5.9e-17;
Matches 285; Conservative 0; Mismatches 269; Indels 15; Gaps 2;

QY 534 GTGCTGCTGCTCTCTCTCTGAGATCCTCTTTGTGGACAGCTTTCCGGCAG 593
Db 827 GTGTATGTGTACACAGTGGCGCTGGCTCTCTCTCTGAGTACCTTTCAACAG 886
QY 594 CACCCGAGTGTCTTACCTGATGAGCCTGCTGGCAGTGTGATGACTTTCACAG 653
Db 887 AACCTGAGGTGTCTCTCTATGAGCCTGTGTGACAGTGTGGCAAAAGTGTACCC 946

QY 654 AGCAGCCTGAGAGCTGACATGAGCTGTGGGAGTCTTCTGCTCCGTCTCTGTGT 713
Db 947 GGGAGAGCGCTTCCCTGAGGGGAGAGCGGAGCATGTGTGAGCCCTCTCCCTG 1006
QY 714 GATATGAGGTCTTTATGCTATGATGAACCCAGGCCCCGGAAAGTCC-----764
Db 1007 GATCTTCGGTTCAGCTGATATACCCCGCAGGAGTGGGGGGGCAACCTCACCAGT 1066
QY 765 ---AGCCTCTTCAGTGGAGCAAAAGCCGGCCCTGTGCTCAGCGCTGTGTACTTC 821
Db 1067 CTGGGATCTTTTGGGAGCAGCAGCTAACAAAGTGTGTGTCTCTGCGCACTCTCTGCC 1126
QY 822 TTCCCTGCCACAGATCAGCTCACCAGCAGCTGCAAGCTGCTC---TGGGTACAGCA 878
Db 1127 TACCGCAAGAGAGTCTCGAGTGTGTGAGACAGCCGGGTGTGCAAAAAGTGGCCACCTCAA 1186
QY 879 CCCTTGATATAGTGTGAGAAAGGCTCCCTCTCAGAGGCTTGTGTGTACTCAAGAGGTG 938
Db 1187 CGCTGGACAGCTTCGAGAGAGGTGTCCAGATACCGCAGCGGTGTATCAAGGGCGTG 1246
QY 939 CGTTTCTGAGCTCGAGGCGCCCTCTATCCACTACTCAGGAGCCCTTCCCTCAACTGCAC 998
Db 1247 CGGGCTTGTGATGTGGCTGTGTGGGCGCCGCTCTTAAGATATCAGCCTTGGACCTCAAG 1306
QY 999 GTGCTGACCTGTGTCCGAGAGCCCGGGCGGCTGTCCGATCCGGAGCACACCCACCA 1058
Db 1307 GTATCCACCTAGTATAGTATCTCTGCTGTGCGAGCTCCCGCATCCGCTCGCTCAC 1366
QY 1059 GAACCTATGTTGACAGTATATTTGCT 1087
Db 1367 GGCCTATCCGGGAAAGCTTACAGGTGT 1395

RESULT 11
US-09-040-984-63
Sequence 63, Application US/09040984
Patent No. 6210883

GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Wang, Tongtong
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS
TITLE OF INVENTION: OF LUNG CANCER
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040,984
FILING DATE: 18-MAR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.456
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-282-6031
TELEX:
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 731 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-09-040-984-63

Query Match 2.7%: Score 52.4; DB 3: Length 731;
Best Local Similarity 57.2%; Pred. No. 5.3e-06;
Matches 95; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 922 TGGTACTCAAGAGAGCGTTCCTTCTCAGCCCTGACGCCCTTATCCACTACTACAGGACC 981
D 3 TAGTCATAAAGGGGTGTGGCGCTTCGACGTGGGGGTCTTGGCGCCACTGCTGGGAGACC 62
QY 982 CTTCCCTCAGCTCAGCTGCTGACCTGCTCGAGAGACCCCGGCGCGTTCGATCCC 1041
D 63 CGGCCCTGAGCTCAGGCTCCTCCATCTGCTGCTGATCCCGCGCGGCGAGTTCAC 122
QY 1042 GGGAGCAGCACCATGAGACTGATGCTGATGATGATGATGCT 1087
D 123 GGATCCCTCGCGCCAGCGGCTCATCCGCTGAGAGCCCTAGAGTGT 168

RESULT 12

US-09-123-912-63

Sequence 63: Application US/09123912A
Patent No. 6312695

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Mang, Tonglong

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER

FILE REFERENCE: 210121.455C1

CURRENT APPLICATION NUMBER: US/09/123.912A

PRIOR FILING DATE: 1998-07-27

PRIOR APPLICATION NUMBER: 09/040.802

NUMBER OF SEQ ID NOS: 114

SOFTWARE: Patentln Ver. 2.0

SEQ ID NO 63

LENGTH: 731

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: modified_base

LOCATION: (236)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (249)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (263)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (288)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (312)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (317)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (326)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (337)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (352)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (362)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (611)

OTHER INFORMATION: where n is a, c, g or t

LOCATION: (370)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (376)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (400)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (411)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (414)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (434)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (436)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (446)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (457)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (473)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (486)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (497)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (498)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (502)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (512)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (531)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (546)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (554)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (563)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (565)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (588)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (597)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (608)

OTHER INFORMATION: where n is a, c, g or t

NAME/KEY: modified_base

LOCATION: (611)

```
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (613)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (615)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (627)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (632)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (640)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (641)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (644)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (654)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (660)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (663)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (665)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (671)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (678)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (692)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (697)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (698)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (699)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (704)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (705)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (712)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (714)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (717)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (718)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (719)
OTHER INFORMATION: Where n is a, c, g or t
```

```
NAME/KEY: modified_base
LOCATION: (723)
Query Match
Best Local Similarity 57.2%; Score 52.4; DB 4; Length 731;
Matches 95; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 922 TGGTACTCAAGAGAGGTGCGCTTTTCACGCTCGAGGCCCTCATCTACTACTACAGGACC 981
DB 3 TAGTCATAAAGGAGGTGCGCGCTTCGACGTGGCGGCTTTGGCGCAGCTGCGAGACC 62
QY 982 CTTCCTCACTGACGCTGTCGACCTGTGCGAGACCCCGGCGCGGTTCATCC 1041
DB 63 CGGCCCTGACCTCAAGGTCATCCACTGTGCTGCTGATCCCGCGGCGAGGTAC 122
QY 1042 GGGAGCACACCACTAGACTCATGTGTGACAGTCATATGTGCT 1087
DB 123 GGATCCGCTCGCGCCACGCGCTCATCCGTGAGAGCCTACAGGTGCT 168
```

```
RESULT 13
US-09-643-597-63
Sequence 63, Application US/09643597
Patent No. 6426072
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Fan, Liqun
APPLICANT: Kalos, Michael D.
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Hosken, Nancy
APPLICANT: Fanger, Gary R.
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aljun
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Henderson, Robert A.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.45311
CURRENT APPLICATION NUMBER: US/09/643,597
NUMBER OF SEQ ID NOS: 369
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 63
LENGTH: 731
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(731)
OTHER INFORMATION: n = A,T,C or G
US-09-643-597-63
```

```
Query Match
Best Local Similarity 57.2%; Score 52.4; DB 4; Length 731;
Matches 95; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 922 TGGTACTCAAGAGAGGTGCGCTTTTCACGCTCGAGGCCCTCATCTACTACTACAGGACC 981
DB 3 TAGTCATAAAGGAGGTGCGCGCTTCGACGTGGCGGCTTTGGCGCAGCTGCGAGACC 62
QY 982 CTTCCTCACTGACGCTGTCGACCTGTGCGAGACCCCGGCGCGGTTCATCC 1041
DB 63 CGGCCCTGACCTCAAGGTCATCCACTGTGCTGCTGATCCCGCGGCGAGGTAC 122
QY 1042 GGGAGCACACCACTAGACTCATGTGTGACAGTCATATGTGCT 1087
DB 123 GGATCCGCTCGCGCCACGCGCTCATCCGTGAGAGCCTACAGGTGCT 168
```

```
RESULT 14
US-09-480-884A-63
Sequence 63, Application US/09480884A
```

Query Match	2.7%	Score	52.4	DB	4	Length	731
Best Local Similarity	57.2%	Pred. No.	5.3e-06				
Matches	95	Conservative	0	Mismatches	71	Indels	0
						Gaps	0

```
Search completed: August 14, 2003, 17:44:32
Job time : 128.227 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:40:04 ; Search time 34.5891 Seconds
(without alignments)
2894.677 Million cell updates/sec

Title: US-09-645-078-4
Perfect score: 2042
Sequence: 1 MMLKRGRLMFLGSGVIVV.....NLSDLLSSHIILQGVFREG 388

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_protent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriopl:*
- 17: sp_archaeopl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2042	100.0	388	11 Q9WUE5	Q9WUE5 mus musculu
2	2033	99.6	388	11 Q9RII1	Q9RII1 mus musculu
3	1513.5	74.1	386	4 Q9Y5R3	Q9Y5R3 homo sapien
4	1510.5	74.0	386	4 Q8NCG5	Q8NCG5 homo sapien
5	1470.5	72.0	370	4 Q8IV46	Q8IV46 homo sapien
6	984	48.2	395	11 Q9OUP4	Q9OUP4 mus musculu
7	977	47.8	395	4 Q8GZS9	Q8GZS9 mus sapien
8	958.5	46.9	411	4 Q8GZS9	Q8GZS9 mus sapien
9	595.5	29.2	484	11 Q9EP78	Q9EP78 mus musculu
10	593	29.0	486	4 Q75667	Q75667 homo sapien
11	592.5	29.0	484	11 Q99NB0	Q99NB0 mus musculu
12	592	29.0	486	4 Q9NSR4	Q9NSR4 homo sapien
13	579.5	28.4	530	11 Q88276	Q88276 mus musculu
14	576.5	28.2	483	4 Q9UED5	Q9UED5 homo sapien
15	576.5	28.2	530	4 Q9Y4C5	Q9Y4C5 homo sapien
16	527.5	25.8	411	4 Q43916	Q43916 homo sapien

17	527.5	25.8	411	11 Q9EQC0	Q9EQC0 mus musculu
18	497	24.3	479	4 Q75099	Q75099 homo sapien
19	478.5	23.4	472	11 Q88199	Q88199 mus musculu
20	454.5	22.3	474	11 Q9QZL2	Q9QZL2 rattus norv
21	399	19.5	441	13 Q93403	Q93403 torpedo cal
22	281	13.8	486	5 Q9VMC3	Q9VMC3 drosophila
23	210.5	10.3	363	5 Q9VMC4	Q9VMC4 drosophila
24	203	9.9	120	11 Q9YDK5	Q9YDK5 mus musculu
25	199.5	9.8	363	5 Q8WZD1	Q8WZD1 drosophila
26	174	8.5	119	6 Q95J48	Q95J48 oryctolagus
27	128	6.3	274	16 Q8PRA0	Q8PRA0 xanthomonas
28	126	6.2	1222	4 Q8IZU8	Q8IZU8 homo sapien
29	124	6.1	307	16 Q92VG4	Q92VG4 rhizobium m
30	98.5	4.8	3187	11 Q63714	Q63714 rattus norv
31	98	4.8	913	2 Q9KID4	Q9KID4 streptomyce
32	97	4.8	303	16 Q93JDE	Q93JDE streptomyce
33	96.5	4.7	379	3 Q8NI23	Q8NI23 neurospora
34	96.5	4.7	834	5 Q9U7C0	Q9U7C0 eutoliiculi
35	96.5	4.7	985	2 P94324	P94324 bradyrhizob
36	96	4.7	876	12 Q8AYX3	Q8AYX3 porcine lym
37	95	4.7	285	2 Q8KRX0	Q8KRX0 xanthomonas
38	95	4.7	388	2 Q87614	Q87614 pseudomonas
39	94.5	4.6	506	8 Q8MCS2	Q8MCS2 galega orle
40	94.5	4.6	513	16 Q9LA28	Q9LA28 escherichia
41	93.5	4.6	522	16 Q8CXZ3	Q8CXZ3 escherichia
42	93	4.6	346	10 Q9FZ80	Q9FZ80 arabidopsis
43	93	4.6	726	16 Q9X839	Q9X839 streptomyce
44	93	4.6	865	12 Q913V4	Q913V4 porcine gam
45	93	4.6	876	12 Q8JYD3	Q8JYD3 porcine lym

ALIGNMENTS

RESULT 1

ID	Q9WUE5	PRELIMINARY:	PRT:	388 AA.
AC	Q9WUE5;			
DT	01-NOV-1999 (TREMBLrel. 12, Created)			
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DE	N-acetylglucosamine 6-O-sulfotransferase (Carbohydrate 6/keratan) sulfotransferase 4).			
GN	CHST4.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6; TISSUE=Tonsil;			
RX	MEDLINE=99264336; PubMed=10330415;			
RA	Bistrup A., Bhakta S., Lee J.-K., Belov Y.Y., Gunn M.D., Zuo F.-R.,			
RA	Huang C.-C., Kannagi R., Rosen S.D., Hemmerich S.;			
RT	"Sulfotransferases of two specificities function in the reconstitution			
RT	of high endothelial cell ligands for L-selectin.";			
RL	J Cell Biol. 145:899-910(1999).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Tongue;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai T., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Resole G., Quackenbush J.,			
RA	Schmitt L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Barsch G.,			
RA	Blake J., Botfeill D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,			

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Noridone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whittaker C., Wilmberg L.,
 RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.,
 RT Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL: AF131236; AAD3016.1; -
 DR EMBL: AK091133; BAB26078.1; -
 DR MGD: MGI:1349479; Chsta4.
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 KW Transferase.
 SQ SEQUENCE 388 AA; 44694 MW; D1E9D7796DF8574D CRC64;

Query Match 100.0%; Score 2042; DB 11; Length 388;
 Best Local Similarity 100.0%; Pred. No. 5.7e-189;
 Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMLKKGRLLMFGSQVIYVAFIHMSVHRHLSQRESRPPVAVLVSSWRSSSPVQOL 60
 DB 1 MMLKKGRLLMFGSQVIYVAFIHMSVHRHLSQRESRPPVAVLVSSWRSSSPVQOL 60
 QY 61 FCGHPDVFYLMPEPAHVMWMTFTSSSTAMKLMMAVRDLRSVFLCDMSVFDAVNPGRKOS 120
 DB 61 FCGHPDVFYLMPEPAHVMWMTFTSSSTAMKLMMAVRDLRSVFLCDMSVFDAVNPGRKOS 120
 QY 121 SLEWQESRALCAPCDEFPFAHEISSPKHCKLLCGQOPDWEKCRSHGFVLEVR 180
 DB 121 SLEWQESRALCAPCDEFPFAHEISSPKHCKLLCGQOPDWEKCRSHGFVLEVR 180
 QY 181 LSLQALYPLTDPSLNHVHLVRDPAVRSREHTTIELMVDSHIVLGOHLETKEEQD 240
 DB 181 LSLQALYPLTDPSLNHVHLVRDPAVRSREHTTIELMVDSHIVLGOHLETKEEQD 240
 QY 241 PYRAMKTIICKSQVDYKAIOTLPEALQOQRYLFRLYEDVRAPLAOTTRLYKFGDLPLH 300
 DB 241 PYRAMKTIICKSQVDYKAIOTLPEALQOQRYLFRLYEDVRAPLAOTTRLYKFGDLPLH 300
 QY 301 LOTWVYNTVRGKMGQAHFTNARNALNVSQAWRSLPYEKYSQLODAGCEANDLGLYQ 360
 DB 301 LOTWVYNTVRGKMGQAHFTNARNALNVSQAWRSLPYEKYSQLODAGCEANDLGLYQ 360
 QY 361 VRSQOEGNLSLDLSSSHILGOVFRG 388
 DB 361 VRSQOEGNLSLDLSSSHILGOVFRG 388

RESULT 2

Q9RL11 PRELIMINARY; PRT; 388 AA.
 ID Q9RL11;
 AC Q9RL11;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE L-selectin ligand sulfotransferase.
 GN CHST4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99361934; PubMed=10435581;
 RA Hirooka N., Petryniak B., Nakayama J., Tsuboi S., Suzuki M., Yeh J.C.,
 RA Izawa D., Tanaka T., Miyasaka M., Lowe J.B., Fukuda M.,
 RT "A novel, high endothelial venule-specific sulfotransferase expresses
 RT 6-sulfo sialyl Lewis(x), an L-selectin ligand displayed by CD34.";
 RL Immunology 11:79-89(1999).
 DR EMBL: AF109155; AAD45579.1; -
 DR MGD: MGI:1349479; Chsta4.
 DR InterPro: IPR000863; Sulfotransferase.

DR Pfam: PF00685; Sulfotransfer; 1.
 KW Lectin; Selectin; Transferase.
 SQ SEQUENCE 388 AA; 44636 MW; 6D53771AFB6884AE CRC64;
 Query Match 99.6%; Score 2033; DB 11; Length 388;
 Best Local Similarity 99.5%; Pred. No. 4.3e-188;
 Matches 386; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMLKKGRLLMFGSQVIYVAFIHMSVHRHLSQRESRPPVAVLVSSWRSSSPVQOL 60
 DB 1 MMLKKGRLLMFGSQVIYVAFIHMSVHRHLSQRESRPPVAVLVSSWRSSSPVQOL 60
 QY 61 FCGHPDVFYLMPEPAHVMWMTFTSSSTAMKLMMAVRDLRSVFLCDMSVFDAVNPGRKOS 120
 DB 61 FCGHPDVFYLMPEPAHVMWMTFTSSSTAMKLMMAVRDLRSVFLCDMSVFDAVNPGRKOS 120
 QY 121 SLEWQESRALCAPCDEFPFAHEISSPKHCKLLCGQOPDWEKCRSHGFVLEVR 180
 DB 121 SLEWQESRALCAPCDEFPFAHEISSPKHCKLLCGQOPDWEKCRSHGFVLEVR 180
 QY 181 LSLQALYPLTDPSLNHVHLVRDPAVRSREHTTIELMVDSHIVLGOHLETKEEQD 240
 DB 181 LSLQALYPLTDPSLNHVHLVRDPAVRSREHTTIELMVDSHIVLGOHLETKEEQD 240
 QY 241 PYRAMKTIICKSQVDYKAIOTLPEALQOQRYLFRLYEDVRAPLAOTTRLYKFGDLPLH 300
 DB 241 PYRAMKTIICKSQVDYKAIOTLPEALQOQRYLFRLYEDVRAPLAOTTRLYKFGDLPLH 300
 QY 301 LOTWVYNTVRGKMGQAHFTNARNALNVSQAWRSLPYEKYSQLODAGCEANDLGLYQ 360
 DB 301 LOTWVYNTVRGKMGQAHFTNARNALNVSQAWRSLPYEKYSQLODAGCEANDLGLYQ 360
 QY 361 VRSQOEGNLSLDLSSSHILGOVFRG 388
 DB 361 VRSQOEGNLSLDLSSSHILGOVFRG 388

RESULT 3

Q9Y5R3 PRELIMINARY; PRT; 386 AA.
 ID Q9Y5R3;
 AC Q9Y5R3;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE N-acetylglucosamine 6-O-sulfotransferase (L-selectin ligand
 DE sulfotransferase GST-3).
 GN GST3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99264336; PubMed=10330415;
 RA Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,
 RA Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.,
 RT "Sulfotransferases of two specificities function in the reconstitution
 RT of high endothelial cell ligands for L-selectin.";
 RL J. Cell Biol. 145:899-910(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21332592; PubMed=11439191;
 RA Yeh J.-C., Hirooka N., Petryniak B., Nakayama J., Ellies L.G.,

RA Rabuka D., Hindsaul O., Marth J.D., Lowe J.B., Fukuda M.;
 RT "Novel sulfated lymphocyte homing receptors and their control by a
 RT corelation between lymf.3-N-acetylglucosaminyltransferase.";
 RL Cell 105:957-969(2001).
 DR EMBL: AF131235; AAC33015.1;
 DR EMBL: AF280088; AAC48246.1;
 DR EMBL: AF149783; AAC48417.1;
 DR InterPro: IPR001092; HLH_baslc.
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00665; Sulfotransferase.
 DR PROSITE: PS00036; HLH_1; 1.
 DR Lectin: Selectin; Transferrase.
 KW SEQUENCE 386 AA; 45133 MW; 0C3BB4022417143A CRC64;

Query Match 74.1%; Score 1513.5; DB 4; Length 386;
 Best Local Similarity 72.8%; Pred. No. 7/7e-138;
 Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;

QY 1 MMLKGRLLMFLGQYIVVALFTIMSVHR--HLSOREESRRPVHVLVLSMRSGSSFFVG 58
 1 MLPLKMKKLLFLVSQAILALFFHMYSHNISLSMKAQPER-MHVLVLSMRSGSSFFVG 59
 DB 59 QLEGOHDPVYLMERPAHVMWMTTSSSTAKMLHMAVRDLRSVFLCDMSVFDAVMNPPGRK 118
 60 QLEGOHDPVYLMERPAHVMWMTTSSSTAKMLHMAVRDLRSVFLCDMSVFDAVMNPPGRK 119
 QY 119 QSSLFQWEOSSALCSAPVCDPEFPAHETSSPKHCKLGGQOPEDVWEKACRSHGVYLVKEV 178
 120 QSSLFQWEOSSALCSAPVCDPEFPAHETSSPKHCKLGGQOPEDVWEKACRSHGVYLVKEV 179
 DB 120 QSSLFQWEOSSALCSAPVCDPEFPAHETSSPKHCKLGGQOPEDVWEKACRSHGVYLVKEV 179
 QY 179 RFLSLQALYPLDPSLNLHVLVLRDPAVRSRREHTTIELMVDSHIVLGQHLLETKEE 238
 180 RFLSLQALYPLDPSLNLHVLVLRDPAVRSRREHTTIELMVDSHIVLGQHLLETKEE 239
 DB 180 RFLSLQALYPLDPSLNLHVLVLRDPAVRSRREHTTIELMVDSHIVLGQHLLETKEE 239
 QY 239 DQPYAKKTIICKSQVDYKAIQTLPEALQRYLFLREDVLRAPLAQTRIKYKVGDFL 298
 240 DQPYAKKTIICKSQVDYKAIQTLPEALQRYLFLREDVLRAPLAQTRIKYKVGDFL 299
 DB 240 DQPYAKKTIICKSQVDYKAIQTLPEALQRYLFLREDVLRAPLAQTRIKYKVGDFL 299
 QY 299 PHLOTWYVNTYRGKMGQAHFHTNARNALNYSQAMRSLPEYKXSQLODACEAMDILGY 358
 300 PHLOTWYVNTYRGKMGQAHFHTNARNALNYSQAMRSLPEYKXSQLODACEAMDILGY 359
 DB 300 PHLOTWYVNTYRGKMGQAHFHTNARNALNYSQAMRSLPEYKXSQLODACEAMDILGY 359
 QY 359 LQVRSQOEGNLSIDLSSHIIGQV 384
 360 RHVRSQOEGNLSIDLSSHIIGQV 385
 DB 360 RHVRSQOEGNLSIDLSSHIIGQV 385

RESULT 4

Q8NCG5 Q8NCG5 PRELIMINARY; PRT; 386 AA.
 AC Q8NCG5;
 DT 01-OCT-2002 (TREMblrel. 22, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Hypothetical protein FLJ90265.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Isogal T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
 RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
 RA Mesubo Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
 RA Hattori A., Okumura K., Iwayanagi T., Ninoiwa K.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK074746; BAC11177.1;
 DR Genew: HGNC:1972; CHST4.
 KW Hypothetical protein.
 SQ SEQUENCE 386 AA; 45160 MW; 861869348319E42A CRC64;

Query Match 74.0%; Score 1510.5; DB 4; Length 386;
 Best Local Similarity 72.5%; Pred. No. 1.5e-137;
 Matches 280; Conservative 50; Mismatches 53; Indels 3; Gaps 2;

QY 1 MMLKGRLLMFLGQYIVVALFTIMSVHR--HLSOREESRRPVHVLVLSMRSGSSFFVG 58
 1 MLPLKMKKLLFLVSQAILALFFHMYSHNISLSMKAQPER-MHVLVLSMRSGSSFFVG 59
 DB 59 QLEGOHDPVYLMERPAHVMWMTTSSSTAKMLHMAVRDLRSVFLCDMSVFDAVMNPPGRK 118
 60 QLEGOHDPVYLMERPAHVMWMTTSSSTAKMLHMAVRDLRSVFLCDMSVFDAVMNPPGRK 119
 QY 119 QSSLFQWEOSSALCSAPVCDPEFPAHETSSPKHCKLGGQOPEDVWEKACRSHGVYLVKEV 178
 120 QSSLFQWEOSSALCSAPVCDPEFPAHETSSPKHCKLGGQOPEDVWEKACRSHGVYLVKEV 179
 DB 120 QSSLFQWEOSSALCSAPVCDPEFPAHETSSPKHCKLGGQOPEDVWEKACRSHGVYLVKEV 179
 QY 179 RFLSLQALYPLDPSLNLHVLVLRDPAVRSRREHTTIELMVDSHIVLGQHLLETKEE 238
 180 RFLSLQALYPLDPSLNLHVLVLRDPAVRSRREHTTIELMVDSHIVLGQHLLETKEE 239
 DB 180 RFLSLQALYPLDPSLNLHVLVLRDPAVRSRREHTTIELMVDSHIVLGQHLLETKEE 239
 QY 239 DQPYAKKTIICKSQVDYKAIQTLPEALQRYLFLREDVLRAPLAQTRIKYKVGDFL 298
 240 DQPYAKKTIICKSQVDYKAIQTLPEALQRYLFLREDVLRAPLAQTRIKYKVGDFL 299
 DB 240 DQPYAKKTIICKSQVDYKAIQTLPEALQRYLFLREDVLRAPLAQTRIKYKVGDFL 299
 QY 299 PHLOTWYVNTYRGKMGQAHFHTNARNALNYSQAMRSLPEYKXSQLODACEAMDILGY 358
 300 PHLOTWYVNTYRGKMGQAHFHTNARNALNYSQAMRSLPEYKXSQLODACEAMDILGY 359
 DB 300 PHLOTWYVNTYRGKMGQAHFHTNARNALNYSQAMRSLPEYKXSQLODACEAMDILGY 359
 QY 359 LQVRSQOEGNLSIDLSSHIIGQV 384
 360 RHVRSQOEGNLSIDLSSHIIGQV 385
 DB 360 RHVRSQOEGNLSIDLSSHIIGQV 385

RESULT 5

Q8IV46 Q8IV46 PRELIMINARY; PRT; 370 AA.
 AC Q8IV46;
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Similar to carbonyldehydrate (N-acetylglucosamine 6-O) sulfotransferase
 4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tissue-Brain;
 RA Strausberg R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC05282; AAH35282.1;
 KW Transferase.
 SQ SEQUENCE 370 AA; 43249 MW; EA8D76EA4E73C625 CRC64;

Query Match 72.0%; Score 1470.5; DB 4; Length 370;
 Best Local Similarity 73.6%; Pred. No. 1e-133;
 Matches 271; Conservative 45; Mismatches 49; Indels 3; Gaps 2;

QY 19 VVALFTHMSVHR--HLSOREESRRPVHVLVLSMRSGSSFFVGOLFQHPDVYLMERPAH 76
 3 ILALFFHMYSHNISLSMKAQPER-MHVLVLSMRSGSSFFVGOLFQHPDVYLMERPAH 61
 DB 3 ILALFFHMYSHNISLSMKAQPER-MHVLVLSMRSGSSFFVGOLFQHPDVYLMERPAH 61
 QY 77 VMNFTSTANKMLHMAVRDLRSVFLCDMSVFDAVMNPPGRKSSLPQWEOSSALCSAPV 136
 62 VMNFTSTANKMLHMAVRDLRSVFLCDMSVFDAVMNPPGRKSSLPQWEOSSALCSAPV 121
 DB 62 VMNFTSTANKMLHMAVRDLRSVFLCDMSVFDAVMNPPGRKSSLPQWEOSSALCSAPV 121
 QY 137 CDFPAREISSPKCKLCKLGGQOPEDVWEKACRSHGVYLVKEVFLSLOALYPLDPSLN 196
 122 CDFPAREISSPKCKLCKLGGQOPEDVWEKACRSHGVYLVKEVFLSLOALYPLDPSLN 181
 DB 122 CDFPAREISSPKCKLCKLGGQOPEDVWEKACRSHGVYLVKEVFLSLOALYPLDPSLN 181
 QY 197 LHVVLVLRDPAVRSRREHTTIELMVDSHIVLGQHLLETKEEDPYAKKTIICKSQVDY 256
 197 LHVVLVLRDPAVRSRREHTTIELMVDSHIVLGQHLLETKEEDPYAKKTIICKSQVDY 256

```

DB 182 LHIYHLVADPRAVFRSRTKGDMLIDSRIVMGHOKIKKEDOPYYWMQVYCSOLETY 241
OY 257 KAIOTLPBALQOQRTFLRYEDLVAPLAQOTRLRYKFGVGLDFPHLOTWYVNTTRGKMGQ 316
DB 242 KTIGSLFPALQERLYLVREEDLAPARVAQTSRMVEFGVLEDFPHLOTWVHNTTRGKMGD 301
OY 317 HAFHTNARNALNVSQAWRMSLPEYKVSQLODAGCAGMMLDGLQVRSQOEGCNLSLDLS 376
DB 302 HAFHTNADALNVSQAWRMSLPEYKVSRLQKACGAMMLDGLRVRSRQOERHLLDLS 361
OY 377 SSHILGOV 384
DB 362 TWYVEQI 369

RESULT 6
OQ0P4 PRELIMINARY; PRT; 395 AA.
OQ0P4:
OY 01-MAY-2000 (TREMBLrel. 13, Created)
OY 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
OY 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
OY N-acetylglicosamine 6-O-sulfotransferase.
OY CHST5 OR I-GLCNAC-6-ST.
OY Mus musculus (Mouse).
OY Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OY Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OY NCBI_TaxID=10090;
OY [1]
OY SEQUENCE FROM N.A.
OY STRAIN=C57BL/6; TISSUE=Intestine;
OY MEDLINE=99423499; PubMed=10491328;
OY Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.,
OY "Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-
OY Sulfotransferase that is Highly Restricted to Intestinal Tissue.";
OY Biochem. Biophys. Res. Commun. 263:543-549(1999).
OY EMBL; AF176841; AAD56003.1; -.
OY EMBL; AF176840; AAD56002.1; -.
OY MGI; MGI:1931825; Chst5.
OY InterPro; IPR000863; Sulfotransferase.
OY Pfam; PF00685; Sulfotransfer; 1.
OY Transferrase.
OY SEQUENCE 395 AA; 44537 MW; 3FDF71E43ED383BE CRC64;

Query Match 48.2%; Score 984; DB 11; Length 395;
Best Local Similarity 54.9%; Pred. No. 1.3e-86;
Matches 207; Conservative 44; Mismatches 114; Indels 12; Gaps 6;

OY 8 RLMEFGSOVIYVALFIHMSVHRHLSQRESRRP-----VHVLVLSWSSGSSFGOLF 61
DB 2 RLPRF-SSTVMSLLMVGTLIVLVSRQVSPSPAGLGERVHVLVLSWSSGSSFGOLF 60
OY 62 GQHPDVFYLMPEAMHVMWTFSTAMKLMAYRDLRSVFLCDMSVFPAYNMPGRKQSS 121
DB 61 SOHPDVFYLMPEAMHVMWTFSTAMKLMAYRDLRSVFLCDMSVFPAYNMPGRKQSS 119
OY 122 LFOHQSAALCSAPVCDFFPAHEISSPKHCKLLCGQPFDMWEKCRSHGVYLVKEVRL 181
DB 120 LFOHVASRALCSPPVCEAFARGNISSEYVCKRPLCATRFGLAQECSSSHVYLVKEVRF 179
OY 182 SLQALYPLTDPSSLVHVLVADPRAVFRSREHTTELMVDSHVLVCOHETKEEDOP 241
DB 180 NLQVYPLTDPALRLVHVLVADPRAVFRSREHTAKALANDNGVILGN-CTWEADPR 238
OY 242 YVAMKICKSOVDIVK-AIQTLPEALQOQRTFLRYEDLVAPLAQOTRLRYKFGVGLDFPH 300
DB 239 LRVNVEGCRSHVRIEALHKKRPFLQDRYLRVREEDLADROLVYIRELYATGTGLP 298
OY 301 LOTWYVNTTRGKMG--QHAFTNARNALNVSQAWRMSLPEYKVSQLODAGCAGMMLDGLY 358
DB 299 LOTWVHNTTRGSGPARREAFKTSRDALSVQAWRHLLPFAKIRRVQELCGALQILGY 358
OY 359 LQVRSQOEGCNLSLDL 375

```

```

DB 359 RSVHSELQORDLSLDL 375

RESULT 7
OQ6ZX3 PRELIMINARY; PRT; 395 AA.
OQ6ZX3:
OY 01-MAR-2001 (TREMBLrel. 16, Created)
OY 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
OY 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
OY N-acetylglicosamine 6-O-sulfotransferase GST-4beta (Corneal
OY N-acetylglicosamine 6-O-sulfotransferase).
OY GST4BETA OR CHST6.
OY Homo sapiens (Human).
OY Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OY Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OY NCBI_TaxID=9606;
OY [1]
OY SEQUENCE FROM N.A.
OY Akama T.O., Nishida K., Nakayama J., Watanabe H., Fujiwara T.,
OY Nakamura T., Doi A., Kawasaki S., Inoue T., Maeda N., Yamamoto S.,
OY Ozaki K., Kinoshita S., Shimomura Y., Tanigami A., Fukuda M.N.;
OY "Macular corneal dystrophy type I and type II are caused by distinct
OY mutations in a new sulphotransferase gene.";
OY Nat. Genet. 26:237-241(2000).
OY EMBL; AF280086; AAG48244.1; -.
OY EMBL; AF218990; AAG26325.1; -.
OY EMBL; AF219991; AAG26327.1; -.
OY GeneW; HGNC:6938; CHST6.
OY InterPro; IPR000863; Sulfotransferase.
OY Pfam; PF00685; Sulfotransfer; 1.
OY Transferrase.
OY SEQUENCE 395 AA; 44098 MW; 433CA60248A46F7 CRC64;

Query Match 47.8%; Score 977; DB 4; Length 395;
Best Local Similarity 53.2%; Pred. No. 6.3e-86;
Matches 207; Conservative 43; Mismatches 97; Indels 42; Gaps 7;

OY 2 MLTKGRILMEFGSOVIYVALFIHMSVHRHLSQRESRRP-----VHVLVLSW 50
DB 14 LLLAQTFLLFLVLS-----RPGSSFPAGGARVHVLVLSW 49
OY 51 RSGSSFGVGLQFGHPDVFYLMPEAMHVMWTFSTAMKLMAYRDLRSVFLCDMSVFPDA 110
DB 50 RSGSSFGVGLQFGHPDVFYLMPEAMHVMWTFSTAMKLMAYRDLRSVFLCDMSVFPDA 109
OY 111 YNMPGRKQSSLFOHQSAALCSAPVCDFFPAHEISSPKHCKLLCGQPFDMWEKCRSH 170
DB 110 YL-PMRRNLSDFQWAVASRALCSPPACSAFPGALISSEAVCKRPLAROSFTLARACRSY 168
OY 171 GVVYLVKEVRFSLQALYPLTDPSSLVHVLVADPRAVFRSREHTTELMVDSHVLVGO 230
DB 169 SHVYLVKEVRFSLQALYPLTDPSSLVHVLVADPRAVFRSREHTAKALANDNGVILGT 228
OY 231 HLETTKEEDOPYYAKKICKSOVDIVKAIQTL-PEALQOQRTFLRYEDLVAPLAQOTTR 288
DB 229 N-CTWEADPRGLRVNVEGCRSHVRIEAL-ATLKKRPFLRGRLRVREEDLAPLEIR 286
OY 289 LYKFGVGLDFPHLOTWYVNTTRGKMG--QHAFTNARNALNVSQAWRMSLPEYKVSQLO 346
DB 287 LYAFGLSLTLPQLEAMVHNTTRGSGPARREAFKTSRNALNVSQAWRHLLPFAKIRRVQ 346
OY 347 DAGCAGMMLDGLQVRSQOEGCNLSLDL 375

```


Db 347 ELCAGALQLGYPYSEDEQRNLADLV 375

RESULT 8

09GZS9 PRELIMINARY; PRT; 411 AA.
 AC 09GZS9; Q9UBV3;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Internal GLICAC-6-sulfotransferase (intestinal N-acetylglucosamine-6-O-sulfotransferase) (N-acetylglucosamine 6-O-sulfotransferase).
 GN CHST5 OR I-GLCNAC-6-ST.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20472330; PubMed=11017086;
 RA Akama T.O., Nishida K., Nakayama J., Matanabe H., Ozaki K., Nakamura T., Dots A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S., Fujiwara T., Thonar E.J., Shimomura Y., Kinoshita S., Tanigami A., Fukuda M.N.;
 RT "Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene.";
 RL Nat. Genet. 26:237-241(2000).
 RN [2]
 RP SEQUENCE OF 22-411 FROM N.A.
 RC TISSUE=Intestine;
 RX MEDLINE=99423499; PubMed=10491328;
 RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerlich S.;
 RT "Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-Sulfotransferase that is Highly Restricted to Intestinal Tissue.";
 RL Biochem. Biophys. Res. Commun. 263:543-549(1999).
 DR EMBL: AF246718; AAG28023.1; -
 DR EMBL: AF219991; AAG26326.1; -
 DR EMBL: AF176839; AAD56001.1; -
 DR EMBL: AF176838; AAD56000.1; -
 DR Gene: HGNC:1973; CHST5.
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 KW Transference.
 SQ SEQUENCE 411 AA; 46160 MW; 97642D54BE926E06 CRC64;
 Query Match 46.98; Score 958.5; DB 4; Length 411;
 Best Local Similarity 55.68; Pred. No. 4.1e-84;
 Matches 193; Conservative 42; Mismatches 87; Indels 25; Gaps 6;
 42 VHVLYDSNRSGSSFFGOLFGHPDVEYLMPEAMHVMPTSTAMKLMHVAVDLLRSVF 101
 63 VHVLYDSNRSGSSFFGOLFGHPDVEYLMPEAMHVMPTSTAMKLMHVAVDLLRSVF 122
 102 LCDMSVFDAYMNGPKRQSSLFQWQESRALSAPVCDFFPAHRISSPKCKLKGQOPFD 161
 123 LCDMSVFDAYM-POSNNLSAFEMWATSRALCSPPACSAFPRGITSQDVCKITCTQPF 161
 162 MVEKARSHGFVYLKAEVRLSLALPLTDPCLNHVHVLVDPRAVRSRHTTTELM 221
 182 LAREACRSYSHVLLKEVREFNLQVPLSLDPALEIRIVHVLVDPRAVRSRHTTTELM 241
 222 VDSHYLVG-----OHLETKEDQPYAKKTIICKSQVDYVKAIOQL-PEALQQR 270
 242 RDNGIYLVGNGKVEADPHILRIE-----VCRSHVRIAEAA-ATLKPPPLNGRY 290
 271 LFLRYEDLVRAPLAQTTRLYKFEVDELPHILQTVYVNVFGKMGQ--NAFHTNANALN 328
 291 RLVRFEEDLAREPLAEIRALYAFGLTLTQLLEAMINHTIGSGIGRIEAFHSSNRAN 350
 329 VSGAMRWSPLEKVSQLODACGAMDLGLYQVRSOQEGNSLDLL 375
 351 VSGAMRWALPFTKILRVOEVCAGALQLGYPYSDQQRDLTLDLV 397

RESULT 9

09EP78 PRELIMINARY; PRT; 484 AA.
 AC 09EP78;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE N-acetylglucosamine 6-O-sulfotransferase GST-5 (N-acetylglucosamine-6-O-sulfotransferase) (2600013M07RIK protein) (Carbohydrate
 DE (N-acetylglucosamine) sulfotransferase 7).
 GN CHST7 OR GST5 OR 2600013M07RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6;
 RX PubMed=10956661;
 RA Bhakta S., Barts A., Bowman K.G., Kao W.M., Polsky I., Lee J.-K., Cook B.N., Bruenl R.E., Rosen S.D., Bertozzi C.R., Hemmerlich S.;
 RT "Sulfation of N-Acetylglucosamine by Chondroitin 6-Sulfotransferase 2 (GST-5).";
 RL J. Biol. Chem. 275:40226-40234(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Uchimura K., Muramatsu T.;
 RT "Identification and molecular characterization of a cDNA encoding a novel N-acetylglucosamine-6-O-sulfotransferase.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Embryo;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S., Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fieischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J., Schirrl L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barab G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seyg T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohsaki S., Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF280089; AAG48247.1; -
 DR EMBL: AB040710; BAB13769.1; -
 DR EMBL: AK011202; BAB27465.1; -
 DR EMBL: BC019204; AAI19204.1; -
 DR MGD: MGI:1891767; Chst7.
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 KW Transference.
 SQ SEQUENCE 484 AA; 54766 MW; 9B195537D7AB7193 CRC64;
 Query Match 29.28; Score 595.5; DB 11; Length 484;
 Best Local Similarity 38.48; Pred. No. 6.7e-49;

[illegible]

Query Match	Best Local Similarity	Matches	Score	DB 4	Length	Indels	Gaps
29.0%	37.7%	51	136	54	10		
38	SRRPVHLVLSWRSQSSFGVQLFGQNDHVVYLIMPRANHVMTFTSSATMKLHMAVRDLL	97					
98	SREKHQHYVHATWRTGSSFFLELFGHQHDVFLYLRPMNHIMQALRGDAESLQGLRDL	157					
98	RSVFLCDMSVDAIYNPG-PRKQ-----SSLFQEQSFAALCSAPYCDFFPAHE---	144					
158	SLRRCDSVSLRYLARPDPAPARADTANLTAALEFRMTKNTKVISCPICGAPAPARAIV	217					
145	ISSPHEKLLCGQOPFDWMERKACSHGFFVYLKEVRFSLQALYPLLDPSLNLHVHLY	203					

Db 218 GLVEETACERSCPPAIALALEACECKRYPVVYINKVRLLDGLVLPVLLRBDGLNKKVOLF 277

Qy 204 RDPRAVERSREHTTELMVDS-----HIVIGOH-----LETIKEE 238

Db 278 RDPRAVHNSRLTSCROGLLEESIOVLRTRGRDGRFHRVLLLAHGCAPGCGSRALPAPRA 337

Qy 239 DOPYA---AMKICKISOVDYVKAIOITLPALOOBYLFLEFYEDYVAPLQOTPLVYKFGV 295

Db 338 D--FELTGALEYICAMLRDLFLPAGCAPRWMLRRIRLRLEEDLYKQPAQLRRLLRFSG 395

Qy 296 DFLPLHQTYVYVNTREGKNG--QHAFTNNARNALNVSQAWRWSLPEYKVSQOLDACEAMD 354

Db 396 RALAALDFALMLMTGCAAYGADRPHLISARDAREAHVNRERLSRQOVQVEACAPAMR 455

Qy 355 LIGYLVQVRSQOE-----QGNLSIDL 374

Db 456 LLAY--PRSGEEDGAEQPRRGETPLDM 480

Query Match	Best Local Similarity	Matches 136;	Conservative 50;	Mismatches 130;	Indels 41;	Gaps 6;
43	HLVLSRSSSSFFGOLFEGHPDVEYLMERPAHYMTFTSTAKLHMAYVDLRSYFL	101	HIYHATKRTGSSFFGELFNQHPDVFYLYDPWMHLMQALYFGDADASTLGALMDKLSLFR	160		
103	CDMSVFADYMPNG-PRKO-----SSLFQWESRALCSAPVCDFFPAHE-----ISSP	161	CDSVLRIRYAPGDDGERAPDSANLTJTMLEFRMRNTKVCSPPLCPAPRAPADGVIED	220		
149	KHCKLLCGQPPDMEYKAKRSHGEVYLVKEVRFSLQALYPLITDPSLHVVHLYVDPRA	221	KACESTCPVSLRALEACRKYPPVYVINDVRLDLGVLVPLLRDPGLNLKVVQLFRDPRA	280		
209	VFRSEHTITELMDSHIVL-----GQULETI-----KEDDQY	281	VHNSLTKRQGLRESIGVLTQRGDFHRYLLAHGVDAFEGGARALPSAPRADDFLT	340		
243	YAMKIKTSOYDIYVAIOTLPBALQOYLFLEEDLVAPRLAQOTRYLKFEVLDLPHIQ	341	SALEYICAWMLRDLFTGTGAPAWLRKRLRLRIEDLVQPOAQQLRRLRFSGLRTIALID	400		
303	TMVYNVTRGKMG- -OAHFTNARNALNYSQAWMSLPYEKVSQLODACGEAMDILGY	401	AFAPMTGSAVGADRPFLHSARDRAEVAHMRRELSQEQVROQVETACAPARLLAY	457		

Query Match.	29.08;	Score 592;	DB 4;	Length 486;
Best Local Similarity	37.78;	Pred. No. 1.5e-48;		
Matches 146;	Conservative 51;	Mismatches 136;	Indels 54;	Gaps 10;

RESULT 13
088276
ID 088276
PRELIMINARY;
PRT; 530 AA.

OX	NCHI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE-Embryo;
RX	MEDLINE=98380482; PubMed=9712885;
RA	Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurogawa N.,
RA	Mitsukawa C., Kanmagi R., Habuchi O., Muramatsu T.;
RT	"Molecular cloning and characterization of an N-acetylglucosamine-6-O-
RL	sulfotransferase";
RU	J. Biol. Chem. 273:22577-22583(1998).
DR	EMBL; AB011452; BAA32138.1; -
DR	EMBL; AB011452; BAA32138.1; -
DR	EMBL; AB011451; BAA32137.1; -
MGD	MGDI:1891160; Chst2.
DR	InterPro: IPR000863; Sulfotransferase.
DR	Pfam; PF00685; Sulfotransfer; 1.
KW	transferase.
SQ	SEQUENCE 530 AA; 57814 MW; A113EB735C363EC CRC64;
Query Match	28.4%; Score 579.5; DB 11; Length 530;
Best Local Similarity	35.4%; Pred. No. 2,7e-47;
Matches 132;	Conservative 64; Mismatches 122; Indels 55; Gaps 9
OY	44 VLVLSWRSGSSFWGQLFGHDPDYFYLEMPAHVMVMTFTSTAMKLNHAVRDLSRVLC 103
Dd	167 YVFVFTTWSRSGSSFGEELFNQNPEVEFFLYEPWHVQVKLYPGDAVSIGGAARDMLSATLRC 226
OY	104 DMSVPDAVMNGPKRKQS----SLQWEGSRALCSAPVDGFPAHE-----ISSPKHKLL 154
Dd	227 DLVSFQFLSPAGSGGRNLTTGLGFAATNKVCSSPLC--PAYKEVVGLDVRCK-K 282
OY	155 CGOQPFDMVEAKCRSHGFVLKLEAFELSLAFLPDLTPDSLNVHLVLRDPRAVRSRE 214
Dd	283 CPQRALARFECECRKYRTVIKGVAVFDVAVALPLKDPALDKVIHILVRDPAVASR- 341
OY	215 HTTELMWDIVVLGHNEHTKEEP-----OPY-----AM 245
Dd	342 -----IRSRGILLRESLQVVRSRDPRAHRMFEEAAGHKLGAKKEGGMRADYHALGM 395
OY	246 KIICKSQVDIYKAIOITL---PEALQORYLFLREDDLVAPRLAQTRTLRYKGGLDFPLHQ 302
Dd	396 EVICNS---MAKTLOTALQPPDWLQGHVLLVREDDVDDPYKTLRRYVDGFLGLSPENK 452
OY	303 TWVTYNVTRKGKMGCHAEFTNARNALINVSQAARMSLPYEKVSOLODACBANDLGLLOYR 362
Dd	453 QFALNMPTSGSSSSSKPFVVASARNATAQANAMWTALTFOOI KOVEEFCYQPAVLGYERYN 512
OY	363 SQEOGNLSIDL 375
Dd	513 SPEEVKDLSTLL 525
RESULT 14	
OYUED5	
ID	OYUED5 PRELIMINARY; PRT; 483 AA.
AC	OYUED5-
DT	01-MAY-2000 (TREMBLrel. 13, Created)
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE	N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST).
GN	GNCST.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE-Brain;
RX	MEDLINE=98391845; PubMed=9722682;
RA	Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
RA	Mitsukawa C., Kanmagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA	Ozaki T., Nakagawara A., Kadomatsu K., Muramatsu T.;
RT	"Human N-acetylglucosamine-6-O-sulfotransferase involved in the

RT biosynthesis of 6-sulfo steryl Lewis X: Molecular cloning, chromosomal
 mapping, and expression in various organs and tumor cells.";
 RT J. Biochem. 124:670-678(1998).
 DR EMBL; AB014679; BAA34265.2;
 KW Transferase.
 SO SEQUENCE 483 AA; 52787 MW; E27797D44931BA18 CRC64;
 Query Match 28.2%; Score 576.5; DB 4; Length 483;
 Best Local Similarity 35.1%; Pred. No. 4,6e-47;
 Matches 131; Conservative 65; Mismatches 122; Indels 55; Gaps 9;
 QY 44 VVLSWRSGSSFFGOLFGQHPDVFYLMRPANHYMTFTSSSTAKMLHMAVRDLRSVFLC 103
 DB 120 VYFTTWRSGSSFFGELRNQNPFFLEPVMWQKLYPGDAVSLQGAARMLSLALYRC 179
 QY 104 DMSVFDAYNMPGRKOS----SLFQWESRALCSAPVCDFFPAHE-----ISSPKCKLL 154
 DB 180 DLSVFLYSPAGSGGRNLTTLGIFGAATKKVVCSSPLC---PAYREVGVLVDRCK-K 235
 QY 155 CGQGFPMYKAKRSHGFYVLEKVRFLSLQALYPLTDPSSLNHHVHLVDRPRAVRSRE 214
 DB 236 CPQRLARFEEDCRKRYRTLVIGRVFVDAVLAFLRLDPAIDLKVIHLVDRPRAVASSR- 294
 QY 215 HTTIELMVDSHVIGOHLETFKEED-----OPY-----AM 245
 DB 295 -----ISRHGLIRESLQVNSRDPRAHMPFLAAGHKLGAKKEGCGPADYHALGAM 348
 QY 246 KIICKSQVDIVAIQTL--PEALQORYLFLEYEDLVRAPLAQTTRLYKFGVGLDFLPHQ 302
 DB 349 EVICNS---MAKTLQALQPPDMLQGHYLVREEDLVGDPVTKLRVYDFVGLVSPENE 405
 QY 303 TWYVNTYRGKGQHAFTNARNALNVSAQMRWSLPYEKVSQLODAGSAMDLGLYQVR 362
 DB 406 QPALNMTSSGSSSKPFVVSARNATQANAMWTLTFFQIKOVEECYOPMAVLGYERYN 465
 QY 363 SQOEGCNLSLDLL 375
 DB 466 SPEEVKDLSTLL 478
 RESULT 15
 QYAC5 PRELIMINARY; PRT; 530 AA.
 AC QYAC5; OGGZNS; QY6P2;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Long form of N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST)
 DE (Carbohydrate sulfotransferase 2).
 GN GN6ST OR CHST2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RX MEDLINE=98391845; PubMed=9722682;
 RA Uchimura K., Muramatsu H., Kanane T., Ogawa H., Yamakawa T., Fan Q.,
 RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
 RA Ozaki T., Nakagawa A., Kadomatsu K., Muramatsu T.,
 RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
 RT biosynthesis of 6-sulfo steryl Lewis X: Molecular cloning, chromosomal
 RT mapping, and expression in various organs and tumor cells.";
 RL J. Biochem. 124:670-678(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-umbilical vein endothelial cells;
 RX MEDLINE=99168906; PubMed=10049591;
 RA Li X., Tedder T.F.;
 RT "CHST1 and CHST2 sulfotransferases expressed by human vascular
 RT endothelial cells: cDNA cloning, expression, and chromosomal
 RT localization.";

RL Genomics 55:345-347(1999).
 RN [3]
 RP SEQUENCE OF 48-530 FROM N.A.
 RC TISSUE-Placenta;
 RX MEDLINE=20498786; PubMed=11042394;
 RA Sakaguchi H., Kitagawa H., Sugahara K.;
 RT "Functional expression and genomic structure of human N-
 RT acetylglucosamine-6-O-sulfotransferase that transfers sulfate to b-N-
 RT acetylglucosamine at the nonreducing end of an N-acetylglucosamine
 RT sequence.";
 RL Biochim. Biophys. Acta 1523:269-276(2000).
 DR EMBL; AB014680; BAA34266.2;
 DR EMBL; AF083066; ABD20981.1;
 DR EMBL; AB021125; BAB16887.1;
 DR EMBL; AB021124; BAB16886.1;
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 KW Transferase.
 SO SEQUENCE 530 AA; 57857 MW; A82CA227BD5651B CRC64;
 Query Match 28.2%; Score 576.5; DB 4; Length 530;
 Best Local Similarity 35.1%; Pred. No. 5,2e-47;
 Matches 131; Conservative 65; Mismatches 122; Indels 55; Gaps 9;
 QY 44 VVLSWRSGSSFFGOLFGQHPDVFYLMRPANHYMTFTSSSTAKMLHMAVRDLRSVFLC 103
 DB 167 VYFTTWRSGSSFFGELRNQNPFFLEPVMWQKLYPGDAVSLQGAARMLSLALYRC 226
 QY 104 DMSVFDAYNMPGRKOS----SLFQWESRALCSAPVCDFFPAHE-----ISSPKCKLL 154
 DB 227 DLSVFLYSPAGSGGRNLTTLGIFGAATKKVVCSSPLC---PAYREVGVLVDRCK-K 282
 QY 155 CGQGFPMYKAKRSHGFYVLEKVRFLSLQALYPLTDPSSLNHHVHLVDRPRAVRSRE 214
 DB 283 CPQRLARFEEDCRKRYRTLVIGRVFVDAVLAFLRLDPAIDLKVIHLVDRPRAVASSR- 341
 QY 215 HTTIELMVDSHVIGOHLETFKEED-----OPY-----AM 245
 DB 342 -----ISRHGLIRESLQVNSRDPRAHMPFLAAGHKLGAKKEGCGPADYHALGAM 395
 QY 246 KIICKSQVDIVAIQTL--PEALQORYLFLEYEDLVRAPLAQTTRLYKFGVGLDFLPHQ 302
 DB 396 EVICNS---MAKTLQALQPPDMLQGHYLVREEDLVGDPVTKLRVYDFVGLVSPENE 452
 QY 303 TWYVNTYRGKGQHAFTNARNALNVSAQMRWSLPYEKVSQLODAGSAMDLGLYQVR 362
 DB 453 QPALNMTSSGSSSKPFVVSARNATQANAMWTLTFFQIKOVEECYOPMAVLGYERYN 512
 QY 363 SQOEGCNLSLDLL 375
 DB 513 SPEEVKDLSTLL 525
 Search completed: August 9, 2003, 16:44:22
 Job time : 36.5891 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:41:19 ; Search time 16.5426 Seconds
(Without alignments) 992.382 Million cell updates/sec

Title: US-09-645-078-4

Perfect score: 2042
Sequence: 1 MLLKKGRLLMFLGSGQIVV.....NLSIDLSSSHILGQVREG 388

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCUOS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1513.5	74.1	386	US-09-045-284A-2	Sequence 2, Appl1
2	1513.5	74.1	386	US-09-190-911-1	Sequence 1, Appl1
3	1448.5	70.9	483	US-09-786-240-11	Sequence 11, Appl1
4	579.5	28.4	483	US-09-263-023-2	Sequence 2, Appl1
5	579.5	28.4	483	US-09-471-867-2	Sequence 2, Appl1
6	573.5	28.1	484	US-09-263-023-4	Sequence 4, Appl1
7	573.5	28.1	484	US-09-471-867-4	Sequence 4, Appl1
8	527.5	25.8	411	US-09-015-188-2	Sequence 2, Appl1
9	497	24.3	479	US-08-899-514-2	Sequence 2, Appl1
10	496.5	24.3	458	US-08-655-878-2	Sequence 2, Appl1
11	93	4.6	865	US-09-612-204B-24	Sequence 24, Appl1
12	91	4.5	932	US-08-887-534A-45	Sequence 45, Appl1
13	91	4.5	932	US-09-527-431-45	Sequence 45, Appl1
14	86	4.2	195	US-09-612-204B-2	Sequence 2, Appl1
15	85.5	4.2	380	US-09-489-847-277	Sequence 277, App
16	85.5	4.2	414	US-09-786-240-9	Sequence 9, Appl1
17	85.5	4.2	415	US-09-489-847-133	Sequence 133, App
18	85.5	4.2	496	US-09-292-768-64	Sequence 64, Appl1
19	83	4.1	283	US-09-609-816-10	Sequence 10, Appl1
20	83	4.1	304	US-09-609-816-5	Sequence 5, Appl1
21	82	4.0	359	US-09-150-133-11	Sequence 11, Appl1
22	82	4.0	359	US-09-150-141-11	Sequence 11, Appl1
23	82	4.0	359	US-09-374-824-11	Sequence 11, Appl1
24	82	4.0	359	US-09-374-824-11	Sequence 11, Appl1
25	82	4.0	359	US-09-374-824-11	Sequence 11, Appl1
26	81	4.0	668	US-09-811-469-6	Sequence 6, Appl1
27	81	4.0	733	US-09-107-532A-5427	Sequence 5427, Ap

28	80	3.9	225	4	US-08-832-488-2	Sequence 2, Appl1
29	79.5	3.9	302	4	US-09-252-991A-22091	Sequence 22091, A
30	79.5	3.9	350	2	US-08-852-481-2	Sequence 2, Appl1
31	79.5	3.9	724	4	US-09-328-352-7710	Sequence 7710, Ap
32	79	3.9	177	4	US-08-469-260A-53	Sequence 53, Appl1
33	79	3.9	177	4	US-08-488-446-53	Sequence 53, Appl1
34	79	3.9	177	4	US-08-467-344A-53	Sequence 53, Appl1
35	79	3.9	522	4	US-09-302-620B-97	Sequence 97, Appl1
36	79	3.9	535	4	US-09-134-001C-3338	Sequence 3338, Ap
37	78.5	3.8	470	2	US-08-959-011-1	Sequence 1, Appl1
38	78.5	3.8	507	2	US-08-845-161A-4	Sequence 4, Appl1
39	78.5	3.8	507	3	US-09-270-751-4	Sequence 4, Appl1
40	78.5	3.8	507	4	US-09-168-218B-5	Sequence 5, Appl1
41	78	3.8	436	2	US-08-576-626A-47	Sequence 47, Appl1
42	77.5	3.8	304	4	US-09-609-816-4	Sequence 4, Appl1
43	77.5	3.8	512	4	US-09-302-620B-103	Sequence 103, App
44	77	3.8	265	4	US-09-328-352-4347	Sequence 4347, Ap
45	77	3.8	607	4	US-09-252-991A-21640	Sequence 21640, A

ALIGNMENTS

RESULT 1									
US-09-045-284A-2									
Sequence 2, Application US/09045284A									
Patent No. 6265192									
GENERAL INFORMATION:									
APPLICANT: Bistrup, Annette									
APPLICANT: Rosen, Steven D.									
APPLICANT: Hemmerich, Stefan									
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3									
FILE REFERENCE: 6510-107US1									
CURRENT APPLICATION NUMBER: US/09/045, 284A									
CURRENT FILING DATE: 1998-03-20									
NUMBER OF SEQ ID NOS: 9									
SOFTWARE: FastSeq for Windows Version 4.0									
SEQ ID NO 2									
LENGTH: 386									
TYPE: PRT									
ORGANISM: Homo sapiens									
US-09-045-284A-2									
Query Match									
Best Local Similarity 74.1%; Score 1513.5; DB 3; Length 386;									
Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;									
QY	1	MLLKKGRLLMFLGSGQIVV	HL	SO	RESRRPVHVLVLSMRSGS	SVFG	58		
DB	1	MLPRKKMLLFLVSQMAI	LL	LF	MYSHNLSLSMKQ	APRR	59		
QY	59	QLFGQHPVFLMPRAHVM	NT	FS	TKLHMAVRDL	RL	SV	FL	CDMSVFPAYNPPGRK
DB	60	QLFGQHPVFLMPRAHVM	NT	FS	TKLHMAVRDL	RL	SV	FL	CDMSVFPAYNPPGRK
QY	119	QSSLFQMPQSAICAPV	CD	FE	PAHEISSPKKCL	L	CG	QF	PDWEKACRSHGFVLR
DB	120	QSSLFQMPQSAICAPV	CD	FE	PAHEISSPKKCL	L	CG	QF	PDWEKACRSHGFVLR
QY	179	RFLSQALYPLLT	PS	LN	LAVVHLVDR	PAV	FR	SR	HEHTTIELMVD
DB	180	RFLSQALYPLLT	PS	LN	LAVVHLVDR	PAV	FR	SR	HEHTTIELMVD
QY	239	DQPIYAMTICKSQ	VD	YK	AIQTL	PE	AL	Q	XLPLFRYEDV
DB	240	DQPIYAMTICKSQ	VD	YK	AIQTL	PE	AL	Q	XLPLFRYEDV
QY	299	PHLOTWYVNTV	RG	MG	QHAFHTN	ARN	L	AN	TSOAMRSLPEKY
DB	300	PHLOTWYVNTV	RG	MG	QHAFHTN	ARN	L	AN	TSOAMRSLPEKY
QY	359	LQVRSQDQGNL	SL	DL	SSSHILG	V	384		
DB	359	LQVRSQDQGNL	SL	DL	SSSHILG	V	384		

```

Db          360 RHVRSEQEQRLNLLDLSTWTVPEQI 385

RESULT 2
US-09-190-911-1
; Sequence 1, Application US/09190911
; Patent No. 6365365
; GENERAL INFORMATION:
; APPLICANT: Bistup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSTYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/09/190,911
; CURRENT FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 386
; TYPE: PRT
; ORGANISM: H. sapiens
US-09-190-911-1

Query Match          74.1%; Score 1513.5; DB 4; Length 386;
Best Local Similarity 72.8%; Pred. No. 8,9e-164;
Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;

QY 1 MMLKKGRLLMFLGSOVIYVALFIHMSVHR--HLSQRESRRPVHVLVLSMRSGSFG 58
DB 1 MLLPKMKLLPLVSOAMALLALFFHMYSHINISLSKAPQR-MHVLVLSMRSGSFG 59
QY 59 QLEGGHPDVFYLMPEAMHVMFTSTSTANKLHMAVADLRSVFLCDMSVFDAYMNGPGR 118
DB 60 QLEGGHPDVFYLMPEAMHVMFTSTSTANKLHMAVADLRSVFLCDMSVFDAYMNGPGR 119
QY 119 QSSLFQWEOGRALCSAPVCDFFPAHISPPKCKLIGCOQPDVMEKACRSHGVYVLE 178
DB 120 QSSLFQWEOGRALCSAPVCDFFPAHISPPKCKLIGCOQPDVMEKACRSHGVYVLE 179
QY 179 RFLSLQALYPLLTDPSTLHVHVLVDRPRAVRSREHTTIELMDSHVLQHLETKEE 238
DB 180 RFLSLQALYPLLTDPSTLHVHVLVDRPRAVRSREHTTIELMDSHVLQHLETKEE 239
QY 239 DQPYAMKTIICKSOVDIYKAIQTLPEALQOQRYLFREYEDLVRAPLAQTTRLYKFGVGLDF 298
DB 240 DQPYAMKTIICKSOVDIYKAIQTLPEALQOQRYLFREYEDLVRAPLAQTTRLYKFGVGLDF 299
QY 299 PHLQTVWVNVTRGKGGMGHAFTNARNALNVSOAMRMSLPYEKVSQLODAGEADNLIG 358
DB 300 PHLQTVWVNVTRGKGGMGHAFTNARNALNVSOAMRMSLPYEKVSQLODAGEADNLIG 359
QY 359 LQVRSQEQGNLSIDLSSSHILGOV 384
DB 360 RHVRSEQEQRLNLLDLSTWTVPEQI 385

```

```

; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
; FILE REFERENCE: PF-0592 PCT
; CURRENT APPLICATION NUMBER: US/09/786,240
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 09/186,779; unassigned; 60/133,6
; PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1998-11-04; 1999-05-11
; SOFTWARE: PERL Program
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 11
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6558935 2617407CD1
US-09-786-240-11

Query Match          70.9%; Score 1448.5; DB 4; Length 386;
Best Local Similarity 71.1%; Pred. No. 2,3e-156;
Matches 275; Conservative 48; Mismatches 59; Indels 5; Gaps 4;

QY 1 MMLKKGRLLMFLGSOVIYVALFIHMSVHR--HLSQRESRRPVHVLVLSMRSGSFG 58
DB 1 MLLPKMKLLPLVSOAMALLALFFHMYSHINISLSKAPQR-MHVLVLSMRSGSFG 59
QY 59 QLEGGHPDVFYLMPEAMHVMFTSTSTANKLHMAVADLRSVFLCDMSVFDAYMNGPGR 118
DB 60 QLEGGHPDVFYLMPEAMHVMFTSTSTANKLHMAVADLRSVFLCDMSVFDAYMNGPGR 119
QY 119 QSSLFQWEOGRALCSAPVCDFFPAHISPPKCKLIGCOQPDVMEKACRSHGVYVLE 177
DB 120 QSSLFQWEOGRALCSAPVCDFFPAHISPPKCKLIGCOQPDVMEKACRSHGVYVLE 178
QY 178 VFLSLQALYPLLTDPSTLHVHVLVDRPRAVRSREHTTIELMDSHVLQHLETKEE 237
DB 179 VFLSLQALYPLLTDPSTLHVHVLVDRPRAVRSREHTTIELMDSHVLQHLETKEE 238
QY 238 EDQPYAMKTIICKSOVDIYKAIQTLPEALQOQRYLFREYEDLVRAPLAQTTRLYKFGVGLDF 297
DB 239 EDQPYAMKTIICKSOVDIYKAIQTLPEALQOQRYLFREYEDLVRAPLAQTTRLYKFGVGLDF 298
QY 298 LPHLQTVWVNVTRGKGGMGHAFTNARNALNVSOAMRMSLPYEKVSQLODAGEADNLIG 357
DB 299 LPHLQTVWVNVTRGKGGMGHAFTNARNALNVSOAMRMSLPYEKVSQLODAGEADNLIG 358
QY 358 YLQVRSQEQGNLSIDLSSSHILGOV 384
DB 359 YHVRSEQEQRLNLLDLSTWTVPEQI 385

RESULT 4
US-09-263-023-2
; Sequence 2, Application US/09263023
; Patent No. 6037159
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadamatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/263,023
; CURRENT FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2

```


Db 296 ----LRSHRGLRESLQVVRSDPRHRMPLEAGHKLGAKEGVGPADYHALGAMEV 351
 QY 248 ICKSOVDIVKAIQTL---PEALQQRVLFLEVEDLVRAPIAQTTRLYKFEVGLDFLHLQTM 304
 Db 352 ICNS---WAKTILQTLAOPPDWLOGHYLVRYEDLVGDPEVKTLRRYIDFVGLVSPEDMOF 408
 QY 305 VYNTVRGKMGQOHAHTNARNALNVSOAMRSLPEYKVSQIADACGEAMDLGLYLQVRSO 364
 Db 409 ALNMTSGSGSSSKPFVVSARNAQTQANAMRFTALFQQLKQVEFEFYQPMAYLGYERVNSP 468
 QY 365 QEOGNLSIDL 375
 Db 469 EEWKLSKTL 479

RESULT 7

-09-471-867-4
 Sequence 4, Application US/09471867
 Patent No. 6453289

GENERAL INFORMATION:
 APPLICANT: Uchimura, Kenji
 APPLICANT: Muramatsu, Hideki
 APPLICANT: Kadomatsu, Kenji
 APPLICANT: Kannagi, Reiji
 APPLICANT: Habuchi, Osami
 APPLICANT: Muramatsu, Takashi
 TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE AND
 TITLE OF INVENTION: DNA ENCODING THE SAME
 FILE REFERENCE: TOYAMA1,001AUS
 CURRENT APPLICATION NUMBER: US/09/471,867
 EARLIER FILING DATE: 1999-12-23
 EARLIER APPLICATION NUMBER: US 09/263,023
 EARLIER FILING DATE: 1999-03-05
 EARLIER APPLICATION NUMBER: JP 10-54007
 EARLIER FILING DATE: 1998-03-05
 EARLIER APPLICATION NUMBER: JP 10-177844
 EARLIER FILING DATE: 1998-06-24
 NUMBER OF SEQ ID NOS: 10
 SOFTWARE: FASTSEQ for Windows Version 3.0
 SEQ ID NO 4
 LENGTH: 484
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-471-867-4

Query Match 28.1%; Score 573.5; Db 4; Length 484;
 Best Local Similarity 35.0%; Pred. No. 1.8e-56;
 Matches 130; Conservative 65; Mismatches 121; Indels 55; Gaps 9;

QY 46 VLSSRRSSSFVGLGQHPDVFYLMPEAMHVMWMTFTSTAMKLMHAYRDLRSVFLCDM 105
 Db 123 VETTRSSSGSFGELEFNQNPVEFLYEPYHWMQILYRGDAVSLQGAARDMLSALYRDL 182
 QY 106 SFEDVYMPGRKOS----SLFQWESRALCSAPVCDFFPAHE-----ISSRKHKLGC 156
 Db 183 SVFQYXSPAGSGRMLTLGIFGATNKVCSPLC---PAYRKEVGLVDDRVCK-KCP 238
 QY 157 QOPEDMVEKCRSHGFVYLKEVRLSLQALYPLTDPSSLNHLVHLVDRPRAVRSRHT 216
 Db 239 FQRLARFECECKRYTLVIKGVRFDAVLAFLDPLDLKVIHLVDRPRAVRSR-- 295
 QY 217 TIELMVDHIVLGOHLETKED-----QPY-----AMKI 247
 Db 296 ----LRSHRGLRESLQVVRSDPRHRMPLEAGHKLGAKEGVGPADYHALGAMEV 351
 QY 248 ICKSOVDIVKAIQTL---PEALQQRVLFLEVEDLVRAPIAQTTRLYKFEVGLDFLHLQTM 304
 Db 352 ICNS---WAKTILQTLAOPPDWLOGHYLVRYEDLVGDPEVKTLRRYIDFVGLVSPEDMOF 408
 QY 305 VYNTVRGKMGQOHAHTNARNALNVSOAMRSLPEYKVSQIADACGEAMDLGLYLQVRSO 364
 Db 409 ALNMTSGSGSSSKPFVVSARNAQTQANAMRFTALFQQLKQVEFEFYQPMAYLGYERVNSP 468

QY 365 QEOGNLSIDL 375
 Db 469 EEWKLSKTL 479

RESULT 8

US-09-015-188-2
 Sequence 2, Application US/09015188C
 Patent No. 6393358

GENERAL INFORMATION:
 APPLICANT: Williams, Kevin J
 APPLICANT: Tabas, Ira
 TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
 TITLE OF INVENTION: 6-Sulfotransferase
 FILE REFERENCE: JEFF-0231
 CURRENT APPLICATION NUMBER: US/09/015,188C
 CURRENT FILING DATE: 1998-01-29
 NUMBER OF SEQ ID NOS: 17
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 2
 LENGTH: 411
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-015-188-2

Query Match 25.8%; Score 527.5; Db 4; Length 411;
 Best Local Similarity 35.6%; Pred. No. 2.4e-51;
 Matches 144; Conservative 63; Mismatches 155; Indels 43; Gaps 14;

QY 6 KGRLLMFCS---QYIVVALFIMSVH-----RHLQR--EES-----RRPVHLY 46
 Db 6 KAVLLALLASAIQYTAIRTFKASFHCPCGLAELRLCEESPTFAVNLISRTTHLI 65
 QY 47 LSSRRSSSFVGLGQHPDVFYLMPEAMHVMWMTFTSTAMKLMH--AVRDLR 98
 Db 66 LATTSRSGSFVGLGQHPDVFYLMPEAMHVMWMTFTSTAMKLMH--AVRDLR 125
 QY 99 SVFLCDMSVFDAYNMPGRKOS--LFQWESRALCSAPVCDFFPAHEISSPKHKLGC 155
 Db 126 SLYDCDLFLENYIRPPVNHHTDRIFRGASRVLCSPVCDPPGADLVLEEGDCVRRK 185
 QY 156 GQOPEDMVEKCRSHGFVYLKEVRLSLQALYPLTDPSSLNHLVHLVDRPRAVRSRHT 215
 Db 186 GLNLTVAEACRSHGFVYLKEVRLSLQALYPLTDPSSLNHLVHLVDRPRAVRSRHT 245
 QY 216 TIELMVDHIVLGOHLETKEDQPY-----YAMKIICKSQVDIVKAIQTLPEALQQR 271
 Db 246 T----FRDTYRLMRLMYGTGR--KPYNDVYQLTTCEDPNSYSTGLMRPWLKGYM 298
 QY 272 FLRYEDLVRAPIAQTTRLYKFEVGLDFLHLQTMVYNTVRG-KMGQOHAHTNARNALNV 330
 Db 299 LVRYEDLVRAPIAQTTRLYKFEVGLDFLHLQTMVYNTVRG-KMGQOHAHTNARNALNV 357
 QY 331 QAMRSLPEYKVSQIADACGEAMDLGLYLQVRSQEOGNLSIDL 375
 Db 358 EAMRSLPEYKVSQIADACGEAMDLGLYLQVRSQEOGNLSIDL 402

RESULT 9

US-08-899-514-2
 Sequence 2, Application US/08899514
 Patent No. 5910581

GENERAL INFORMATION:
 APPLICANT: FUKUDA, MASAKAZU
 APPLICANT: HABUCHI, OSAMI
 TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
 TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESS: KNOBBE, MARTENS, OLSON & BEAR, LLP
 STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR


```

: CITY: NEWPORT BEACH
: STATE: CALIFORNIA
: COUNTRY: US
: ZIP: 92660
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/899,514
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: DANIEL E. ALTMAN
: REGISTRATION NUMBER: 34,115
: REFERENCE/DOCKET NUMBER: TOYAMA21.001AUS
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 714 760 0404
: TELEFAX: 714 760 9502
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 479
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
:
: US-08-899-514-2

```

Query Match 24.3%; Score 497; DB 2; Length 479;

Best Local Similarity 33.2%; Pred. No. 9,1e-48; Mismatches 142; Indels 36; Gaps 12;

```

: QY 34 QRESRPRV-----HYLVLSWRSGSSFFVGOLEFGHPDYFLMEPAHYVMFT-----S 83
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
: Db 119 KESEPRPRAVAGRRRHVLMATRTGSSFVGEFFNOGNIFFLEPLMIERTVSPGPG 178
:
: QY 84 STWKLHMAVRDLRSVFLCDMSVFDAVYMPGRKOSLSQWEO--SRALCAPVCDFEP 141
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
: Db 179 ANAGSALVYRDVLEKOLFCDLVLEHFTPLPEDHLTFPMFRGSSRLCEDEPVCNPF- 237
:
: QY 142 AHEISSPKCK-LICGQPFDMVEKACRSHGFVLKEVRFSLQALVPLTDSPLNHV 200
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
: Db 238 VKVFEVYHCKNRCCGLNLTIAEACRKEHAKVRIROLEFLOPLAEDPRDLRLVI 297
:
: QY 201 HLYRDPRAVFRSR-----EHTTLEMDSHIVLGQHLETKEDQPYAMKTIKSOVD 254
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
: Db 298 QLVDRPRAVLASBMVAFAGKTKWKWLDE--GQ--DGLREE--VQRLRGNCES--- 347
:
: QY 255 IYKAIQTL-----PEALQQRVLEFLRYEDLVRAPIAQTRLKFKVGLDPLPHLOTWVYVTR 310
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
: Db 348 -ILSLSELGLRQPAWLRGRTMLVRYEDVARGPLQKAREMPPAGIPLTPQVEEMIQNTQ 406
:
: QY 311 GKMGQHAFTNARNALNVSQAMRWSLPYEKVSQLODAGCAMDLGLQVRSQOEGNLT 370
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
: Db 407 AANDGSIVSTQ-KNSEGEKWRFSMPFLAQVYQAQPCGPARHLEFYKRLARAALATNR 465
:
: QY 371 SLDLL 375
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
: Db 466 SVSIL 470

```

RESULT 10
US-08-655-878-2

```

: Sequence 2, Application US/08655878
: Patent No. 5827713
: GENERAL INFORMATION:
: APPLICANT: FUKUTA, MASAKAZU
: TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
: NUMBER OF SEQUENCES: 9

```

```

: CORRESPONDENCE ADDRESS:
: ADDRESSEE:
: STREET:
: CITY:
: STATE:
: COUNTRY:
: ZIP:
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/655,878
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME:
: REGISTRATION NUMBER:
: REFERENCE/DOCKET NUMBER:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE:
: TELEFAX:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 458
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
:
: US-08-655-878-2

```

Query Match 24.3%; Score 496.5; DB 2; Length 458;

Best Local Similarity 32.7%; Pred. No. 9,7e-48; Mismatches 133; Indels 43; Gaps 10;

```

: QY 37 ESRRPVAVLVLSWRSGSSFFVGOLEFGHPDYFLMEPAHYVMFT-----STWKLHMA 92
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
: Db 110 EPRR--HYLVLMATRTGSSFFVGEFFNOGNIFFLEPLMIERTVTEPGANAVGSALV 167
:
: QY 93 VRDLRSVFLCDMSVFDAVYMPGRKQ--SLFQWESRALCAPVCDFEPA--HEISSPK 149
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
: Db 168 YKVDLQQLLCLDLYLESFSPAEHILTAALFRGSSSHLCEEPVCT--PSLKVREXY 225
:
: QY 150 HCK-LICGQPFDMVEKACRSHGFVLKEVRFSLQALVPLTDSPLNHVHLYRDPRA 208
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
: Db 226 HCKNRCCGPLNLTIAEACRKEHAKVRIROLEFLOPLAEDPRDLRLIOLVRDPRA 285
:
: QY 209 VFRSREHTTELMVDSHIVLGQHLETKEDQPYAMKTIKSOVDIVKAIQ----- 260
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
: Db 286 VLVSF-----NVAESKYESWK-----WAAEGEAPLQEDDEVQRLRGNCESIRL 329
:
: QY 261 -----TLPEALQQRVLEFLRYEDLVRAPIAQTRLKFKVGLDPLPHLOTWVYVTRGKM 314
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
: Db 330 SAELGLNQPAWLRGRTMLVRYEDVARGPLQKAREMPPAGIPLTPQVEEMIRANTQAP-Q 388
:
: QY 315 GQHAFTNARNALNVSQAMRWSLPYEKVSQLODAGCAMDLGLQVRSQOEGNLSLDL 374
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
: Db 389 DSNQIVSTQKNSSEPFKEKWRFSIPFLAQVYQADACEPAMRIFGKLASSAOELTRSLST 448
:
: QY 375 L 375
:   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
: Db 449 L 449

```

RESULT 11
US-09-612-204B-24

```

: Sequence 24, Application US/09612204B
: Patent No. 6461811
: GENERAL INFORMATION:
: APPLICANT: Patience, Clive

```

; TITLE OF INVENTION: Swine Gamma Herpesvirus DNA and Methods of Use
 ; FILE REFERENCE: 61/50-299
 ; CURRENT APPLICATION NUMBER: US/09/612,204B
 ; CURRENT FILING DATE: 2001-08-13
 ; PRIOR APPLICATION NUMBER: U.S. 60/142,736
 ; PRIOR FILING DATE: 1999-07-08
 ; PRIOR APPLICATION NUMBER: U.S. 60/168,532
 ; PRIOR FILING DATE: 1999-12-02
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 24
 ; LENGTH: 865
 ; TYPE: prt
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Deduced amino
 ; OTHER INFORMATION: acid sequence of porcine gamma herpesvirus gpb
 ; OTHER INFORMATION: gene
 ; -09-612-204B-24

Query Match 4.6%; Score 93; DB 4; Length 865;
 Best Local Similarity 23.8%; Pred. No. 0.26;
 Matches 43; Conservative 20; Mismatches 50; Indels 68; Gaps 7;

QY 212 SEHTIELMDSHVIVGQ-----HLETKEDQPYAKTIKCSQVDIVKAI 259
 DB 571 SRPYTFKFMNDSTIKYQGLGVNNEILLTTTYLETCQENTFYF----- 614
 QY 260 QTLPEALQORYFLREYEDLVRAPLAQTRRLYKFGVGLDFPHIQTWVYVY-----TRG 311
 DB 615 ---QAKDMYIKYKYEHLKTYPLSSITTLDFINALNF-----TLENDQVYIELYTRD 665
 QY 312 KGMGHAFTNARNALNSQAMR-WSLPEKYS-----QLDAGCEAMDL 356
 DB 666 E-----KRLSNVFDIETMREYNYAQRVSGLRKDLDLSTNRNCFVDAFGSLMDL 717
 QY 357 G 357
 DB 718 G 718

RESULT 12
 US-08-887-534A-45
 ; Sequence 45, Application US/08887534A
 ; Patent No. 6455323

GENERAL INFORMATION:

APPLICANT: Holden, David W.

TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS

NUMBER OF SEQUENCES: 106

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 233 South Wacker Drive/6300 Sears Tower

CITY: Chicago

STATE: Illinois

COUNTRY: United States of America

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/887,534A

FILING DATE:

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Rin-Laures, Li-Hsien

REGISTRATION NUMBER: 33,547

REFERENCE/DOCKET NUMBER: 28341/33996

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312) 474-6300

TELEFAX: (312) 474-0448

TELEX: (312) 474-6600

; INFORMATION FOR SEQ ID NO: 45:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 932 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-887-534A-45

Query Match 4.5%; Score 91; DB 4; Length 932;
 Best Local Similarity 20.0%; Pred. No. 0.49;
 Matches 70; Conservative 58; Mismatches 112; Indels 110; Gaps 18;

QY 23 FTHMSVYHRLSORESRKRVHYL-----VLSKRSQSSRYGQLGHPDVFYLMERA 74
 DB 621 FSHRAVYLDHEQGTGYTPPLHHVPOKATFDIHNPSLSAAVGYEYGNVE---NKKK 676
 QY 75 WHVWTFSTSTAMKLHMAVRDLRSVFLCDMS--VFDAVNGPKROSSLFQWEDSRAIC 132
 DB 677 FVMEAQYGDFA-----NMSQMTFDNFL-----FSSRSKMGERSGILT 713
 QY 133 SAPVCDFFP-AHEISSPKHCKL-----LCGQDFDVEYKACRSHGFVYLK-EVRFSL 183
 DB 714 L-----FLPHAYEGGQPEHSARLERFLQLAENNCTVYVNLSSSYFHLRAQAASIDS 768
 QY 184 QALYPLDTPSLNLYVHYVPRPRAVRSREHTTEIAMDSHYVIGQHETKEEDQPY 243
 DB 769 EQMRPL-----VMSPKSLRNK---TVAKPID-BFTSGGEPILTESYQADK 812
 QY 244 AKKIC---KSDQVYKAIOTLPE-----ALQORYLFRYBDLVRAPLAQTRRLYKFGV 294
 DB 813 YKVLILATGKMTIDLEAKLNKPNDESVLVALERLXP--PPEETETALAQ----- 861
 QY 295 LDFLPHLQ--TWVYVNTRGKMGQAHFHTNARNALNSQAMRWMSLPEYEV 342
 DB 862 ---LPHLEVSVMWQEEPKNG-----AMLYVYPIYV 890

RESULT 13
 US-09-527-431-45
 ; Sequence 45, Application US/09527431
 ; Patent No. 6485899

GENERAL INFORMATION:

APPLICANT: Holden, David W.

TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS

NUMBER OF SEQUENCES: 106

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 233 South Wacker Drive/6300 Sears Tower

CITY: Chicago

STATE: Illinois

COUNTRY: United States of America

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/527,431

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/887,534

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Rin-Laures, Li-Hsien

REGISTRATION NUMBER: 33,547

REFERENCE/DOCKET NUMBER: 28341/33996

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312) 474-6300

TELEFAX: (312) 474-0448

TELEX: (312) 474-6600

INFORMATION FOR SEQ ID NO: 45:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 932 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-527-431-45

```

```

Query Match          4.5%; Score 91; DB 4; Length 932;
Best Local Similarity 20.0%; Pred. No. 0.49;
Matches 70; Conservative 58; Mismatches 112; Indels 110; Gaps 18;

```

```

QY 23 FHMVSRHRLSQRESRRPVHL-----VLSSRRSSSFYGLFGHPDYLMERA 74
DB 621 FSHRAVHLDEQGTGYPLHHVPOKATFDHNSPLSAAYGFEYGNVE-----NKS 676
QY 75 WYVWHTFSTYMKLHMAVRDLRSVFLDMS--VDAVMNNGPKRSSLFQWEGSRALC 132
DB 677 FHWMEQYGDFA-----NMSQMTFDNFL-----FSSRSKRGERSGLT 713
QY 133 SAPVCDFFP-AHEISSPKHCKL-----LCGQPFDMVEKACRSHGFVVK-EVRFSL 183
DB 714 L-----FLPHAYEGCGPHESSARLERFLDLAENCTVYNLSSSYFHLRAQASLDS 768
Y 184 QALYPLDPSLNLAHVHLVPRPRAVRSREHTTEILAVDSHIVLGQHLETKEDQPY 243
DB 769 EQMRPL-----VYMSPKSLRNK---TVAKPID-EFTSGFEPILTESYQADK 812
QY 244 AMKILC---KSOVDIYKATQTLPE-----ALQGRFLFYREEDLVAPLAQTRLYKFG 294
DB 813 VTKVILATGKMTDLKALKNPDESULLVAIERLPF--PEEIEIALLAQ----- 861
QY 295 LDELPHILQ--TWYVYNTRGKGMGOHAFHTNARNALNVAQMRKSLPEYKY 342
DB 862 ---LPLLEVSVMQVEBKNG-----AMLYVYYVAV 890

```

```

RESULT 14
US-09-612-204B-2
; Sequence 2, Application US/09612204B
; Patent No. 6461811
; GENERAL INFORMATION:
; APPLICANT: Patience, Clive
; TITLE OF INVENTION: Swine Gamma Herpesvirus DNA and Methods of Use
; FILE REFERENCE: 61750-299
; CURRENT APPLICATION NUMBER: US/09/612,204B
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: U.S. 60/142,736
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: U.S. 60/168,532
; PRIOR FILING DATE: 1999-12-02
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Deduced amino
; OTHER INFORMATION: acid sequence derived from the first open reading
; OTHER INFORMATION: frame of the DNA of SED ID NO:1
US-09-612-204B-2

```

```

Query Match          4.2%; Score 86; DB 4; Length 195;
Best Local Similarity 26.5%; Pred. No. 0.16;
Matches 26; Conservative 11; Mismatches 29; Indels 32; Gaps 2;

```

```

QY 212 SREHTTIELMVDSHIVLQ-----HLETKEDQPYAMKIICKSQVDIYKAI 259
DB 88 SRPVTFFKFMNDSTIKYGGVNNELLFTTYLETQCEMTEYF----- 131
QY 260 QTLPEALQRYLFLKREDLVRAPLAOTRLYKFGVLD 297
DB 132 ---QAKTMYIYKNYEHKLTVPLSITLDFIALNF 165

```

```

RESULT 15
US-09-489-847-277
; Sequence 277, Application US/09489847
; Patent No. 6476195
; GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: P2031P1
; CURRENT APPLICATION NUMBER: US/09/489,847
; CURRENT FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/US99/17130
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 60/095,486
; EARLIER FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 60/096,319
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: 60/095,454
; EARLIER FILING DATE: 1998-08-06
; EARLIER APPLICATION NUMBER: 60/095,455
; EARLIER FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 376
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 277
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-489-847-277

```

```

Query Match          4.2%; Score 85.5; DB 4; Length 380;
Best Local Similarity 23.1%; Pred. No. 0.51;
Matches 42; Conservative 24; Mismatches 77; Indels 39; Gaps 6;

```

```

QY 144 EISSPKHCKLGGQPFDMVEKACRSHGFVYKVEVF-LSQALYPLDPSLNLAHVH- 201
DB 198 KVKLKRYTFLEFRDPFVRLISAFRS-----KFELENEDFYKRFVAPMLRLYANH 248
QY 202 -LYRDPRAVRSREHT-----IELAVDSHIVLGQHLETKEDQPYAMKIICKSQVDI 256
DB 249 SLPASAREAFRAKLKVSFANFIQYLDLPH---EKLAFNEHWRQYRLCHPCQIDYDFV 305
QY 257 KATQTLPEALQRYLFLRYEDLV-----APLAQTRLYKFGVLD 296
DB 306 GKLETTDEDAQQLQLQYLDVDRQLRPPPSYRNKRTASSWEDWPAKIFLARQQLKYLEAD 365
QY 297 FL 298
DB 366 FV 367

```

```

Search completed: August 9, 2003, 16:45:38
Job time : 18.5426 secs

```

THIS PAGE RI ANK (11/10/10)


```

OY 179 RFLSLQALYPLLTDPSSLNLHVHLVDRPRAVERSHREHTTIELMDVSHVLGQHELTKEE 238
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 180 RFFNLQSLYPLTKDPSLNLHVHLVDRPRAVERSHREHTTIELMDVSHVLGQHELTKEE 239
OY 239 DOPPYAMITICKSOVDYIKATQTLPEALQOQRLFRYEDYRAPLAQOTRLTKYGLDEL 298
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 240 DOPPYAMOVITQSOLEIKTKTOSLPKALQOERYLVRVBDLAPAPAOYSRMTEFVGLDEL 299
OY 299 PHLOTWVNYTRGKMGQHAFTNARNALNVSQAWRMSLPYKVSQLODACEAMDLLGY 358
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 300 PHLOTWVNYTRGKMGQHAFTNARNALNVSQAWRMSLPYKVSRLKACGDAANLLGY 359
OY 359 LQVNSQOEGNLSIDLSSSHILGOV 384
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 360 RHVRSEQOERNLLDLSTWTVPQEI 385
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

SURF 2
-10-007-262-1
; Sequence 1, Application US/10007262
; Publication No. US20020164748A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/10/007,262
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 386
; TYPE: PRT
; ORGANISM: H. sapiens
US-10-007-262-1

```

```

Query Match 74.1%; Score 1513.5; DB 14; Length 386;
Best Local Similarity 72.8%; Pred. No. 2.9e-151;
Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;

```

```

OY 1 MMLKKGRLMFLGSQVIVVALFTIMSVHR--HLSQREBSRRPVHVLVSSWRSGSSPVG 58
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MLPPKMKLLFLVSQMAIILAFEMYSHNISLSLTKAQPFR--MHVLVLSWRSGSSPVG 59
OY 59 QLEFGHPDVFYILMEPAMHVMTFTSTAMKLMAYRDLIRSVFLCDMSVFDAYMPPGRK 118
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 60 QLEFGHPDVFYILMEPAMHVMTFTSTAMKLMAYRDLIRSVFLCDMSVFDAYMPPGRK 119
OY 119 QSSLEFQWBSRALCSAPVCDFFPAHEISSPKHCKLLCGQPPDMEKACRSHGFVLKEV 178
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 120 QSSLEFQWBSRALCSAPVCDFFPAHEISSPKHCKLLCGQPPDMEKACRSHGFVLKEV 179
OY 179 RFLSLQALYPLLTDPSSLNLHVHLVDRPRAVERSHREHTTIELMDVSHVLGQHELTKEE 238
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 180 RFFNLQSLYPLTKDPSLNLHVHLVDRPRAVERSHREHTTIELMDVSHVLGQHELTKEE 239
OY 239 DOPPYAMITICKSOVDYIKATQTLPEALQOQRLFRYEDYRAPLAQOTRLTKYGLDEL 298
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 240 DOPPYAMOVITQSOLEIKTKTOSLPKALQOERYLVRVBDLAPAPAOYSRMTEFVGLDEL 299
OY 299 PHLOTWVNYTRGKMGQHAFTNARNALNVSQAWRMSLPYKVSQLODACEAMDLLGY 358
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 300 PHLOTWVNYTRGKMGQHAFTNARNALNVSQAWRMSLPYKVSRLKACGDAANLLGY 359
OY 359 LQVNSQOEGNLSIDLSSSHILGOV 384
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 360 RHVRSEQOERNLLDLSTWTVPQEI 385
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 3
US-09-927-602-5
; Sequence 5, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; FILE REFERENCE: P-1J 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-927-602-5

```

```

Query Match 49.0%; Score 1001; DB 9; Length 418;
Best Local Similarity 54.5%; Pred. No. 4.6e-97;
Matches 207; Conservative 45; Mismatches 118; Indels 10; Gaps 4;

```

```

OY 4 LKKGRLMFLGSQVIVVALFTIMSVHR--HLSQREBSRRPVHVLVSSWRSGSSPVG 57
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 21 LARGHRLPFRSSTVHLSLMTVGTGLVFLVSNQVSSPAGLGERHVLVLSWRSGSSPVG 80
OY 58 GOLFGHPDVFYILMEPAMHVMTFTSTAMKLMAYRDLIRSVFLCDMSVFDAYMPPGRK 117
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 81 GOLFGHPDVFYILMEPAMHVMTFTSTAMKLMAYRDLIRSVFLCDMSVFDAYMPPGRK 139
OY 118 KOSLFEQWBSRALCSAPVCDFFPAHEISSPKHCKLLCGQPPDMEKACRSHGFVLKEE 177
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 140 NISDLFQWBSRALCSAPVCDFFPAHEISSPKHCKLLCGQPPDMEKACRSHGFVLKEE 199
OY 178 RFLSLQALYPLLTDPSSLNLHVHLVDRPRAVERSHREHTTIELMDVSHVLGQHELTKEE 237
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 200 VAFNLQVLYPLTDPSSLNLHVHLVDRPRAVERSHREHTTIELMDVSHVLGQHELTKEE 258
OY 238 EQQPYAMITICKSOVDYIKATQTLPEALQOQRLFRYEDYRAPLAQOTRLTKYGLDEL 297
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 239 ADPRLRVNVEGRSHVRLAEALHKKPPFLQDRTYRLVRVEDLARDPLTVIRELYAFTGLGL 318
OY 238 LPHLOTWVNYTRGKMG--QHAFTNARNALNVSQAWRMSLPYKVSQLODACEAMDLL 355
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 319 TQLOLQWVNYTRGKMG--QHAFTNARNALNVSQAWRMSLPYKVSQLODACEAMDLL 378
OY 356 LQVNSQOEGNLSIDLSSSHILGOV 375
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 379 LQVNSQOEGNLSIDLSSSHILGOV 398
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 4
US-09-927-602-3
; Sequence 3, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; FILE REFERENCE: P-1J 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 395

```

TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: synthetic construct
 NAME/KEY: VARIANT
 LOCATION: (1)...(395)
 OTHER INFORMATION: Xaa = any amino acid
 US-09-927-602-3

Query Match 47.9%: Score 978; DB 9; Length 395;
 Best Local Similarity 53.28; Pred. No. 1.1e-94;
 Matches 207; Conservative 42; Mismatches 98; Indels 42; Gaps 7;

QY 2 MLTKKGRLLMFLGSOVIVLALFIMSVHRHLSQREESRRP-----VHVLVLSW 50
 DB 14 LLLAQFTLLLEFVS-----RGPSSPAGEKRVHVLVLSW 49
 QY 51 RSGSSFVGOLFQGHDPVFLMEPAMHVMFTSTANKLHMAVRDLRSVFLCDMSYFDA 110
 DB 50 RSGSSFVGOLFQGHDPVFLMEPAMHVMFTSTANKLHMAVRDLRSVFLCDMSYFDA 109
 QY 111 YNMGPRKOSLSLFWESRALCSAPVCDFFPAHHSPPKCKLKGQOPDMVKACRSH 170
 DB 110 YL-PWRRLSDLFQMAVSRLCSPACSAFPGXISSEYCKPLCARQPFYLAREACRSY 168
 QY 171 GFVVLKEVRFSLQALYPLTDPSLNHLVHLVDRPAVRSRHTTIELVDSHIVLQ 230
 DB 169 SHVVLKEVRFENLQVLPFLSDPALNRIYHLVDRPAVRSRQTKALARDNGIYLG 228
 QY 231 HLETIKEEDOPYAMKTIKCSQVDIVAKIOTL--PEALQORYFLRYEDLVRAPLAQTTR 288
 DB 229 N-GTWVEADPLRVAVREYCSHVRIAE-ATLKPPPLRGYRLVREDLAREPLAEIRA 286
 QY 289 LYKRVGDELPHLOTWYNTYTRGKMG--QHAFTNARNALNVSQARMSLPYKVSQLO 346
 DB 287 LYAFGTGLTPQLAEAMHNTHTGSGPGARREAFKTSRNALNVSQARHMLPFAKIRRVQ 346
 QY 347 DACGAMDILGYLQVRSQOEGNLSIDL 375
 DB 347 ELCAQALQLLGRVYSEDEQRNLALDLV 375

RESULT 5
 US-09-927-602-2
 Sequence 2, Application US/09927602
 Patent No. US20020061562A1
 GENERAL INFORMATION:
 APPLICANT: Fukuda, Michiko N.
 APPLICANT: Akema, Tomoya O.
 TITLE OF INVENTION: Methods of Treating Macular Corneal
 FILE REFERENCE: P-LJ 4852
 CURRENT APPLICATION NUMBER: US/09/927,602
 CURRENT FILING DATE: 2001-08-09
 PRIOR APPLICATION NUMBER: US 09/638,211
 PRIOR FILING DATE: 2000-08-11
 NUMBER OF SEQ ID NOS: 38
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 2
 LENGTH: 395
 TYPE: PRT
 ORGANISM: Homo Sapien
 US-09-927-602-2

Query Match 47.8%: Score 977; DB 9; Length 395;
 Best Local Similarity 53.28; Pred. No. 1.4e-94;
 Matches 207; Conservative 43; Mismatches 97; Indels 42; Gaps 7;
 QY 2 MLTKKGRLLMFLGSOVIVLALFIMSVHRHLSQREESRRP-----VHVLVLSW 50
 DB 14 LLLAQFTLLLEFVS-----RGPSSPAGEKRVHVLVLSW 49
 QY 51 RSGSSFVGOLFQGHDPVFLMEPAMHVMFTSTANKLHMAVRDLRSVFLCDMSYFDA 110

DB 50 RSGSSFVGOLFQGHDPVFLMEPAMHVMFTSTANKLHMAVRDLRSVFLCDMSYFDA 109
 QY 111 YNMGPRKOSLSLFWESRALCSAPVCDFFPAHHSPPKCKLKGQOPDMVKACRSH 170
 DB 110 YL-PWRRLSDLFQMAVSRLCSPACSAFPGXISSEYCKPLCARQPFYLAREACRSY 168
 QY 171 GFVVLKEVRFSLQALYPLTDPSLNHLVHLVDRPAVRSRHTTIELVDSHIVLQ 230
 DB 169 SHVVLKEVRFENLQVLPFLSDPALNRIYHLVDRPAVRSRQTKALARDNGIYLG 228
 QY 231 HLETIKEEDOPYAMKTIKCSQVDIVAKIOTL--PEALQORYFLRYEDLVRAPLAQTTR 288
 DB 229 N-GTWVEADPLRVAVREYCSHVRIAE-ATLKPPPLRGYRLVREDLAREPLAEIRA 286
 QY 289 LYKRVGDELPHLOTWYNTYTRGKMG--QHAFTNARNALNVSQARMSLPYKVSQLO 346
 DB 287 LYAFGTGLTPQLAEAMHNTHTGSGPGARREAFKTSRNALNVSQARHMLPFAKIRRVQ 346
 QY 347 DACGAMDILGYLQVRSQOEGNLSIDL 375
 DB 347 ELCAQALQLLGRVYSEDEQRNLALDLV 375

RESULT 6
 US-09-927-602-4
 Sequence 4, Application US/09927602
 Patent No. US20020061562A1
 GENERAL INFORMATION:
 APPLICANT: Fukuda, Michiko N.
 APPLICANT: Akema, Tomoya O.
 TITLE OF INVENTION: Methods of Treating Macular Corneal
 FILE REFERENCE: P-LJ 4852
 CURRENT APPLICATION NUMBER: US/09/927,602
 CURRENT FILING DATE: 2001-08-09
 PRIOR APPLICATION NUMBER: US 09/638,211
 PRIOR FILING DATE: 2000-08-11
 NUMBER OF SEQ ID NOS: 38
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 4
 LENGTH: 390
 TYPE: PRT
 ORGANISM: Homo Sapien
 US-09-927-602-4

Query Match 46.9%: Score 958.5; DB 9; Length 390;
 Best Local Similarity 55.6%: Pred. No. 1.3e-92;
 Matches 193; Conservative 42; Mismatches 87; Indels 25; Gaps 6;
 QY 42 VHVLVLSMRSGSSFVGOLFQGHDPVFLMEPAMHVMFTSTANKLHMAVRDLRSVFL 101
 DB 42 VHVLVLSMRSGSSFVGOLFQGHDPVFLMEPAMHVMFTSTANKLHMAVRDLRSVFL 101
 QY 102 LCDMSYFDAVYNNPGRKOSLSLFWESRALCSAPVCDFFPAHHSPPKCKLKGQOPDM 161
 DB 102 LCDMSYFDAVYNNPGRKOSLSLFWESRALCSAPVCDFFPAHHSPPKCKLKGQOPDM 160
 QY 162 MVEKACRSHGFRVYLYKEVRFSLQALYPLTDPSLNHLVHLVDRPAVRSRHTTIELM 221
 DB 161 LAREACRSYSHVLYKEVRFENLQVLPFLSDPALNRIYHLVDRPAVRSRHTTIELM 220
 QY 222 VDSHIVG-----QHAFTNARNALNVSQARMSLPYKVSQLO 270
 DB 221 RDNGIYLGTKNCKWEADPHRLRIE-----VCRSHVRIAE-ATLKPPPLRGY 269
 QY 271 LYKRVGDELPHLOTWYNTYTRGKMG--QHAFTNARNALNVSQARMSLPYKVSQLO 328
 DB 270 LYKRVGDELPHLOTWYNTYTRGKMG--QHAFTNARNALNVSQARMSLPYKVSQLO 329
 QY 329 VDSQARMSLPYKVSQLODACGAMDILGYLQVRSQOEGNLSIDL 375
 DB 330 VDSQARMSLPYKVSQLODACGAMDILGYLQVRSQOEGNLSIDL 376

RESULT 7

US-09-927-602-8
Sequence 8, Application US/09927602
Patent No. US20020061562A1

GENERAL INFORMATION:

APPLICANT: Fukuda, Michiko N.
APPLICANT: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
TITLE OF INVENTION: Dystrophy
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 8

LENGTH: 171

TYPE: PRT

ORGANISM: Homo Sapien

US-09-927-602-8

Query Match 31.9%; Score 651.5; DB 9; Length 171;

Best Local Similarity 47.2%; Pred. No. 1,1e-60; Indels 115; Gaps 3;

Matches 134; Conservative 21; Mismatches 14;

QY 32 LSQRESRRPVHVLSSWRSGSFVGLFGQHPDVFYLMPEPMHVMFTSTAMKLM 91

DB 3 LSMKQPER-MHVLVLSMRSGSFVGLFGQHPDVFYLMPEPMHVMFTSTAMKLM 51

QY 92 AVRDLIRSVFLCDMSVFDAVYANPGRKSSLFQWESRALCSAPVCDFFPAHEISSPKHC 151

DB 52 -----KKKCRSTSHVYLKEVREFNLQSLPLKDSPLNHLVLRDPAVR 51

QY 152 KLICGQPFDMYKACRSHGFVYLKEVREFNLQSLPLKDSPLNHLVLRDPAVR 211

DB 52 -----KKKCRSTSHVYLKEVREFNLQSLPLKDSPLNHLVLRDPAVR 99

QY 212 SREHTTIELMVDSHVILGOHLETKEDQPYAMKICKSOVDYKAIQTLPEALQOARYL 271

DB 100 SRETKGKLMDS-----KTISLKLALQERYL 127

QY 272 FLRYEDYRAPLAQTRLYKFGVGLFPLHLOTWYVNVTRGKGMG 315

DB 128 LVRVEDLARAPVAYQTSRMYEFVGLFPLHLOTWYVNVTRGKGMG 171

RESULT 8

US-10-212-933-2
Sequence 2, Application US/10212933
Publication No. US20030008366A1

GENERAL INFORMATION:

APPLICANT: Uchimura, Kenji
APPLICANT: Muramatsu, Hideki
APPLICANT: Kadomatsu, Kenji
APPLICANT: Kadamagi, Reiji
APPLICANT: Habuchi, Osami
APPLICANT: Muramatsu, Takashi
TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-
TITLE OF INVENTION: SULFOTRANSFERASE AND
FILE REFERENCE: TOYAMA1.001AUS
CURRENT APPLICATION NUMBER: US/10/212,933
CURRENT FILING DATE: 2002-08-05
PRIOR APPLICATION NUMBER: US/09/263,023
PRIOR FILING DATE: 1999-03-05
PRIOR APPLICATION NUMBER: JP 10-54007
PRIOR FILING DATE: 1998-03-05
PRIOR APPLICATION NUMBER: JP 10-177844
PRIOR FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 10

SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 483
TYPE: PRT
ORGANISM: Mus musculus
US-10-212-933-2

Query Match 28.4%; Score 579.5; DB 15; Length 483;

Best Local Similarity 35.4%; Pred. No. 1.9e-52; Indels 55; Gaps 9;

Matches 132; Conservative 64; Mismatches 122;

QY 44 VVLSWRSGSFVGLFGQHPDVFYLMPEPMHVMFTSTAMKLMHVRLLSVFLC 103

DB 120 VVFTWRSGSFVGLFGQHPDVFYLMPEPMHVMFTSTAMKLMHVRLLSVFLC 179

QY 104 DMSVFDAYNPPGRKOS-----SLFQWESRALCSAPVCDFFPAHEISSPKCKLL 154

DB 180 DLSVFLVSPAGSGGRNLTGLIFGATKTKVVCSSPLC---PAYRKEVGLVDRCK-K 235

QY 155 CGQQPFDMYKACRSHGFVYLKEVREFNLQSLPLKDSPLNHLVLRDPAVRFRSRE 214

DB 236 CPQRLAREEEDCRKRYRVVIGVYFVDAVLAFLPKDPALDKYIHLVRDPAVASSR- 294

QY 215 HTTIELMVDSHVILGOHLETKEDQPYAMKICKSOVDYKAIQTLPEALQOARYL 245

DB 295 -----INSRHLIRSLQVRSRDPRAHMPFLPAAGKICAKKCGMPADYHALLGM 348

QY 246 KICKSOVDYKAIQTLPEALQOARYLFLRYEDYRAPLAQTRLYKFGVGLFPLHLO 302

DB 349 EVICNS---MAKTLQALQPPDMVGHVYVREDYVGPVATLRVYDFVGLVSPKME 405

QY 303 TWYVNVTRGKGMGQHNFTNARNALNVSOAMWMSLPEYVSOLODAGCAMLLGLGYR 362

DB 406 QFALNMTSGSGSSKRFVYSARNATQANANWRTALTFOIKQVEECYQPMALGYERYN 465

QY 363 SOEOGNLSLDLL 375

DB 466 SPEEVKDLSTLL 478

RESULT 9

US-10-212-933-4
Sequence 4, Application US/10212933
Publication No. US20030008366A1

GENERAL INFORMATION:

APPLICANT: Uchimura, Kenji
APPLICANT: Muramatsu, Hideki
APPLICANT: Kadomatsu, Kenji
APPLICANT: Kadamagi, Reiji
APPLICANT: Habuchi, Osami
APPLICANT: Muramatsu, Takashi
TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-
TITLE OF INVENTION: SULFOTRANSFERASE AND
FILE REFERENCE: TOYAMA1.001AUS
CURRENT APPLICATION NUMBER: US/10/212,933
CURRENT FILING DATE: 2002-08-05
PRIOR APPLICATION NUMBER: US/09/263,023
PRIOR FILING DATE: 1999-03-05
PRIOR APPLICATION NUMBER: JP 10-54007
PRIOR FILING DATE: 1998-03-05
PRIOR APPLICATION NUMBER: JP 10-177844
PRIOR FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 484
TYPE: PRT
ORGANISM: Homo sapiens
US-10-212-933-4

Query Match

28.1%; Score 573.5; DB 15; Length 484;

Best Local Similarity 35.0%; Pred. No. 8.4e-52;

Matches 132; Conservative 64; Mismatches 122;


```

:
:
: CURRENT FILING DATE: 2001-08-09
: PRIOR APPLICATION NUMBER: US 09/638,211
: PRIOR FILING DATE: 2000-08-11
: NUMBER OF SEQ ID NOS: 38
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 7
: LENGTH: 169
: TYPE: PRT
: ORGANISM: Homo Sapien
: OS-09-927-602-7

```

Query Match	23.0%	Score 469.5	DB 9	Length 169
Best Local Similarity	37.6%	Pred. No. 1.8e-41		
Matches 103	Conservative 19	Mismatches 39	Indels 113	Gaps 3

```

QY      42  VHHVLYSSRRSSSEFGQLFGQHPPVFLYLMPEAHVMTTSTAMKLHMAVRDLRSVF 101
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
          9  VHHVLYSSRRSSSEFGQLFGQHPPVFLYLMPEAHVMTT----- 47

QY      102 LCDMSVFDAYMNPGRKSSLFQWQESRALCSAPYCDFEFAHETSSPRCKLLGGQPDF 161
          48 ----- 47

Db      162 MVEKACRSHGFVYLKEVRLSLQALYPLTDPDSLNLHYHLYRDPRAVRSREHTTELM 221
          :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
          48 -LSACRSSTSHVYLKEFRFENLQVLYPLTSDPALRLRYHLYRDPRAVLSREAGPTLA 106
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
          222 VDSHIVLGHLETFIREEDQPYVAMKLTICKSQVDIVKALQTLPEALQOQRYLFLRYEDVRA 281
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
          107 RDNEAA-----TLKP-----PPLRGRYRLVREDIARE 135

QY      282 PLAAQTRLLKFFYGDELPHLTQTVVYNYNRGGMG 315
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
          136 PLAEIRALYAFGLTLTPQLEAMHINITHGSGIG 169

```

```

RESULT 13
US-09-927-602-9
: Sequence 9, Application US/09927602
: Patent No. US20020061562A1
: GENERAL INFORMATION:
: APPLICANT: Fukuda, Michiko N.
: APPLICANT: Akama, Tomoya O.
: TITLE OF INVENTION: Methods of Treating Macular Corneal
: TITLE OF INVENTION: Dystrophy
: FILE REFERENCE: P-1J 4852
: CURRENT APPLICATION NUMBER: US/09/927,602
: CURRENT FILING DATE: 2001-08-09
: PRIOR APPLICATION NUMBER: US 09/638,211
: PRIOR FILING DATE: 2000-08-11
: NUMBER OF SEQ ID NOS: 38
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 9
: LENGTH: 169
: TYPE: PRT
: ORGANISM: Homo Sapien
: US-09-927-602-9

```

Query Match	15.2%	Score 310;	DB 9;	Length 169;
Best Local Similarity	29.1%	Pred. No. 1.3e-24;		
Matches 78;	Conservative 26;	Mismatches 50;	Indels 114;	Gaps 4;

[illegible]

Oy	226	IIVGQHETIKEDQGYAMKTIKCSQVDIVAIOGLEALQGRYLELREDELVRALAQ	285
			: : :
Dd	112	TAL-----QP-----PDMQGHHYLVRRDELDGDPVKI	139
Oy	286	TTPLAKFKGLDPLPHIQITVNVNVTBKG	313
		: : :	
Dd	140	LRRVIDVGVLVSPEMEQPALMNTGGSG	167

```

RESULT 14
US-09-927-602-10
: Sequence 10, Application US/09927602
: Patent No. US20020061562A1
:
GENERAL INFORMATION:
:
APPLICANT: Fukuda, Michiko N.
:
TITLE OF INVENTION: Methods of Treating Macular Corneal
:
TITLE OF INVENTION: Dystrophy
:
FILE REFERENCE: P-LJ 4852
:
CURRENT APPLICATION NUMBER: US/09/927,602
:
CURRENT FILING DATE: 2001-08-09
:
PRIORITY APPLICATION NUMBER: US 09/638,211
:
PRIORITY FILING DATE: 2000-08-11
:
NUMBER OF SEQ ID NOS: 38
:
SOFTWARE: FastSeq for Windows Version 4.0
:
SEQ ID NO 10
:
LENGTH: 179
:
TYPE: PRF
:
ORGANISM: Homo Sapien
:
US-09-927-602-10

```

Query Match	14.1%;	Score 288.5;	DB 9;	Length 179;
Best Local Similarity	27.6%;	Pred. No. 2.6e-22;		
Matches 75;	Conservative 24;	Mismatches 62;	Indels 111;	Gaps 4;

[illegible]

```

RESULT 15
US-09-927-602-11
Sequence 11, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPLICANT: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
Dystrophy
FILE REFERENCE: P-1J 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0

```


THIS PAGE BLANK (USPTO)

THIS PAGE BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:38:24 ; Search time 39.6021 Seconds
(without alignments)
1555.119 million cell updates/sec

Title: US-09-645-078-4
Perfect score: 2042
Sequence: 1 MMLKKRLLMFLFGQIVY.....NLSDLLSSHLIGQVREG 388

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_19Jun03.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2042	100.0	388	20	AAV39919
2	1513.5	74.1	386	20	AAV39918
3	1510.5	74.0	386	22	AAV39918
4	1498.5	73.4	380	23	AAU11274
5	1468.5	70.9	386	21	AAV79219
6	1062	52.0	418	21	AAV81947
7	1001	49.0	418	22	AAV81557
8	984	48.2	395	22	AAV72638
9	984	48.2	395	23	AAV11275

10	978	47.9	395	23	ABB81555
11	977	47.8	395	22	AAV72640
12	977	47.8	395	23	ABB81554
13	977	47.8	395	23	AAE15438
14	958.5	46.9	390	22	AAV72639
15	958.5	46.9	390	22	ABB81556
16	651.5	31.9	171	23	ABB81560
17	592	29.0	486	24	ABP56121
18	579.5	28.4	483	20	AAV31656
19	576.5	28.2	530	22	AAV35367
20	573.5	28.1	484	20	AAV31657
21	573.5	28.1	531	19	AAV69414
22	527.5	25.8	411	19	AAV61100
23	527.5	25.8	411	23	AAE25356
24	527.5	25.8	411	24	ABU03503
25	507	24.8	499	24	ABR11139
26	497	24.3	479	19	AAW02863
27	496.5	24.3	458	18	AAW06480
28	481.5	23.6	169	23	ABB81558
29	469.5	23.0	169	23	ABB81559
30	310	15.2	169	23	ABB81561
31	288.5	14.1	179	23	ABB81562
32	274	13.4	174	23	ABB81563
33	210.5	10.3	363	22	ABB64512
34	176	8.6	183	22	ABB6582
35	134.5	6.6	388	24	AAE32782
36	130	6.4	1207	22	AAV72643
37	130	6.4	1207	24	AAE33542
38	126	6.2	596	22	AAV72641
39	126	6.2	1212	24	AAE33541
40	126	6.2	1222	22	AAV72642
41	126	6.2	1222	24	ABU11849
42	111	5.4	315	22	ABB64513
43	102	5.0	673	22	ABG18753
44	94.5	4.6	524	22	ABB59149
45	93	4.6	346	23	ABB91056

ALIGNMENTS

RESULT 1
ID AAV39919 standard; Protein: 388 AA.
AC AAV39919;
DT 08-DEC-1999 (first entry)
DE Mouse glycosyl sulfotransferase-3 protein sequence.
KW Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
KW selectin binding interaction; inflammation; lymphocyte homing; mouse;
KW secondary lymph organ.
XX Mus. sp.
XX W09949018-A1.
XX 30-SEP-1999.
XX PD 26-FEB-1999; 99WO-US04316.
XX PE 20-MAR-1998; 98US-0045284.
XX PR 12-NOV-1998; 98US-0190911.
XX PA (REGC) UNITEX CALIFORNIA.
XX PA (SYNT) SYNTX USA INC.
XX PI Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
XX WPI: 1999-580442/49.
XX DR N-PSDB; AA220793.

XX Human and murine glycosyl sulfotransferase 3 and related
 PT polynucleotides
 XX
 PS Claim 2; Fig 4; 59pp; English.

CC This sequence is the mouse glycosyl sulfotransferase-3 (GST-3) of
 CC the invention. The nucleic acid sequences, probes and primers derived
 CC from these, proteins and antibodies are useful in detecting homologues.
 CC The sequences, antibodies and methods are useful in the diagnosis and
 CC treatment of diseases associated with selectin binding interactions,
 CC including conditions associated with or resulting from the homing of
 CC leukocytes to sites of inflammation and the normal homing of lymphocytes
 CC to secondary lymph organs.

XX Sequence 388 AA;

Query Match 100.0%; Score 2042; DB 20; Length 388;
 Best Local Similarity 100.0%; Pred. No. 1.2e-208;
 Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMLKKGRLMFLGSGVIVVAFIHMSVHRHLSORESRRPVHVLVLSMRSGSPVGL 60
 DB 1 MMLKKGRLMFLGSGVIVVAFIHMSVHRHLSORESRRPVHVLVLSMRSGSPVGL 60
 QY 61 FQGHDPVFTLMEPRAMHVMFTSTAWKILMAVRDLRSVFLCDMSVFDAYMNPGRKOS 120
 DB 61 FQGHDPVFTLMEPRAMHVMFTSTAWKILMAVRDLRSVFLCDMSVFDAYMNPGRKOS 120
 QY 121 SLFQWESRALCSAPVCDFFPAHEISSPKHCKLGGQOPDMVEKACRSHGFVYLKEVF 180
 DB 121 SLFQWESRALCSAPVCDFFPAHEISSPKHCKLGGQOPDMVEKACRSHGFVYLKEVF 180
 QY 181 LSIQALYPLTDPISLHVHLVDRPRAVRSREHTTIELMVDSHVYLGQHLETKEDQ 240
 DB 181 LSIQALYPLTDPISLHVHLVDRPRAVRSREHTTIELMVDSHVYLGQHLETKEDQ 240
 QY 241 PYYAKTIKSGVDIVKAIQTLPEALQORLFLREEDLVRAPLAQTRLYKFGVGLDFPH 300
 DB 241 PYYAKTIKSGVDIVKAIQTLPEALQORLFLREEDLVRAPLAQTRLYKFGVGLDFPH 300
 QY 301 LQTVVYNTVRGKGQGHAFHTNARNALNVSOAMRWSLPYEKVSQLODACEAMDLGLGYLQ 360
 DB 301 LQTVVYNTVRGKGQGHAFHTNARNALNVSOAMRWSLPYEKVSQLODACEAMDLGLGYLQ 360
 QY 361 VRSQOEGNLSIDLSSSHILGQVFREG 388
 DB 361 VRSQOEGNLSIDLSSSHILGQVFREG 388

RESULT 2
 AAY39918
 ID AAY39918 standard; Protein; 386 AA.

XX AAY39918;
 DT 08-DEC-1999 (first entry)

DE Human glycosyl sulfotransferase-3 protein sequence.

XX glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
 KM selectin binding interaction; inflammation; lymphocyte homing; human;
 KW secondary lymph organ.

XX Homo sapiens.

XX MO9949018-A1.

XX 30-SEP-1999.

XX 26-FEB-1999; 99WO-US04316.

XX 20-MAR-1998; 98US-0045284.

PR 12-NOV-1998; 98US-0190911.
 XX
 PA (REGC) UNIV CALIFORNIA.
 PA (SYNT) SYNTX USA INC.

PI Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
 DR WPI; 1999-580442/49.
 DR N-PSDB; AA220792.

PT Human and murine glycosyl sulfotransferase 3 and related
 PT polynucleotides

PS Claim 2; Fig 1; 59pp; English.

CC This sequence is the human glycosyl sulfotransferase-3 (GST-3) of
 CC the invention. The nucleic acid sequences, probes and primers derived
 CC from these, proteins and antibodies are useful in detecting homologues.
 CC The sequences, antibodies and methods are useful in the diagnosis and
 CC treatment of diseases associated with selectin binding interactions,
 CC including conditions associated with or resulting from the homing of
 CC leukocytes to sites of inflammation and the normal homing of lymphocytes
 CC to secondary lymph organs.

XX Sequence 386 AA;

Query Match 74.1%; Score 1513.5; DB 20; Length 386;
 Best Local Similarity 72.8%; Pred. No. 2.6e-152;
 Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;

QY 1 MMLKKGRLMFLGSGVIVVAFIHMSVHR--HLSORESRRPVHVLVLSMRSGSPVGL 58
 DB 1 MMLKKGRLMFLGSGVIVVAFIHMSVHR--HLSORESRRPVHVLVLSMRSGSPVGL 58
 QY 59 OLFQGHDPVFTLMEPRAMHVMFTSTAWKILMAVRDLRSVFLCDMSVFDAYMNPGRK 118
 DB 59 OLFQGHDPVFTLMEPRAMHVMFTSTAWKILMAVRDLRSVFLCDMSVFDAYMNPGRK 118
 QY 60 QLFQGHDPVFTLMEPRAMHVMFTSTAWKILMAVRDLRSVFLCDMSVFDAYMNPGRK 119
 DB 60 QLFQGHDPVFTLMEPRAMHVMFTSTAWKILMAVRDLRSVFLCDMSVFDAYMNPGRK 119
 QY 119 QSLFQWESRALCSAPVCDFFPAHEISSPKHCKLGGQOPDMVEKACRSHGFVYLKEV 178
 DB 119 QSLFQWESRALCSAPVCDFFPAHEISSPKHCKLGGQOPDMVEKACRSHGFVYLKEV 178
 QY 179 RELSIALYPLTDPISLHVHLVDRPRAVRSREHTTIELMVDSHVYLGQHLETKEDQ 238
 DB 179 RELSIALYPLTDPISLHVHLVDRPRAVRSREHTTIELMVDSHVYLGQHLETKEDQ 238
 QY 239 DQPYAMKTIKSGVDIVKAIQTLPEALQORLFLREEDLVRAPLAQTRLYKFGVGLDFL 298
 DB 239 DQPYAMKTIKSGVDIVKAIQTLPEALQORLFLREEDLVRAPLAQTRLYKFGVGLDFL 298
 QY 240 DQPYAMKTIKSGVDIVKAIQTLPEALQORLFLREEDLVRAPLAQTRLYKFGVGLDFL 299
 DB 240 DQPYAMKTIKSGVDIVKAIQTLPEALQORLFLREEDLVRAPLAQTRLYKFGVGLDFL 299
 QY 299 PHLQTVVYNTVRGKGQGHAFHTNARNALNVSOAMRWSLPYEKVSQLODACEAMDLGLGY 358
 DB 299 PHLQTVVYNTVRGKGQGHAFHTNARNALNVSOAMRWSLPYEKVSQLODACEAMDLGLGY 358
 QY 300 PHLQTVVYNTVRGKGQGHAFHTNARNALNVSOAMRWSLPYEKVSQLODACEAMDLGLGY 359
 DB 300 PHLQTVVYNTVRGKGQGHAFHTNARNALNVSOAMRWSLPYEKVSQLODACEAMDLGLGY 359
 QY 359 LQYRSQOEGNLSIDLSSSHILGQVFREG 384
 DB 359 LQYRSQOEGNLSIDLSSSHILGQVFREG 384
 QY 360 RHVRSEQEQRNLLDLSTWTVPEOI 385
 DB 360 RHVRSEQEQRNLLDLSTWTVPEOI 385

RESULT 3
 AAM93309
 ID AAM93309 standard; Protein; 386 AA.

XX AAM93309;

XX 06-NOV-2001 (first entry)

XX Human polypeptide; SEQ ID NO: 2817.

XX Human; full length cDNA; cDNA synthesis; oligo-capping.

XX Homo sapiens.

PN EP1130094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-0114089.
 XX
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI: 2001-524255/58.
 DR N-PSDB: AAK94229.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 XX
 PS Claim 8; SEQ ID NO 2817; 1380bp + sequence listing; English.
 CC
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a polypeptide
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 CC
 SQ Sequence 386 AA;
 Query Match 74.0%; Score 1510.5; DB 22; Length 386;
 Best Local Similarity 72.5%; Pred. No. 5.4e-152;
 Matches 280; Conservative 50; Mismatches 53; Indels 3; Gaps 2;
 QY 1 MLLKGRLLMFLGQYIVVALFTIMSVHR--HLSQREESRRPVHVLVLSWRSGSSFPVG 58
 DB 1 MLLPKKKKLLFLVSGMAIILALFFHMYSHNLSLSMAQPER-MHVLVLSWRSGSSFPVG 59
 QY 59 QLEQGHVDVFTLMPAHVMTFTSSSTAMKILMAVRLSLSVFLCDMSVPDAYNPGPRK 118
 DB 60 QLEQGHVDVFTLMPAHVMTFTSSSTAMKILMAVRLSLSVFLCDMSVPDAYNPGPRK 119
 QY 119 QSLFQWESRALCSAVPCDFEPFAHETISPKHCKLLCGOOPEDVWEKACRSHGVVLKEV 178
 DB 120 QSLFQWESRALCSAVPCDFEPFAHETISPKHCKLLCGOOPEDVWEKACRSHGVVLKEV 179
 QY 179 RFLSLQALYPLLDPSINLHVHLVDRPRAVFRSREHTTELMVDSHIVLQGHLEETKEE 238
 DB 180 RFLSLQALYPLLDPSINLHVHLVDRPRAVFRSREHTTELMVDSHIVLQGHLEETKEE 239
 QY 239 DQPYAKTIICKSQVDVKAQIOTPEALQRYLEFLREDEVRAVLAQTTRIKYVGDFEL 298
 DB 240 DQPYAKTIICKSQVDVKAQIOTPEALQRYLEFLREDEVRAVLAQTTRIKYVGDFEL 299
 QY 299 PHLTQVYVNTYRGKMGQAHFHTNARNALNVSQAMRSLPEPEKSQLQDAGCEMDLGLY 358
 DB 300 PHLTQVYVNTYRGKMGQAHFHTNARNALNVSQAMRSLPEPEKSQLQDAGCEMDLGLY 359
 QY 359 LQYRSQOEGNLSLDLSSHIILGOV 384
 DB 360 RHVRSQEOQRNLIDLILSTWTVPQOI 385

ID AA011274 standard; Protein; 380 AA.
 XX
 AC AA011274;
 XX
 DT 12-MAR-2002 (first entry)
 XX
 DE Human L-selectin sulfotransferase-2 (LSST-2) protein.
 XX
 KW Human; beta1,3GNT; beta1,3-N-acetylgalactosaminyltransferase; MECA-79;
 KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
 KW ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;
 KW allergic contact dermatitis; lymphoma; chronic pneumonia; LSST-2;
 KW delayed-type hypersensitivity reaction; hyperplastic thymus; antileuker;
 KW antineoplastic; antipsoriatic; antidiabetic; dermatological;
 KW antiallergic.
 KW
 OS Homo sapiens.
 XX
 PN WO200185177-A1.
 XX
 PD 15-NOV-2001.
 XX
 PF 10-MAY-2001; 2001WO-US15452.
 XX
 PR 11-MAY-2000; 2000US-0569320.
 XX
 PA (BURN-) BURINAM INST.
 XX
 PI Fukuda M, Yeh J, Hiraoka N;
 XX
 DR WPI: 2002-075226/10.
 DR N-PSDB: AAS16947.
 XX
 PT New enzyme, useful for modifying acceptor molecule, comprises an
 PT isolated L-selectin sulfotransferase-2 that directs expression of
 PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
 PT intestinal G12MC 6-sulfotransferase -
 PT
 XX
 PS Claim 21; Fig 4; 98pp; English.
 CC
 CC The present invention provides a method of modifying an acceptor molecule
 CC by contacting the acceptor with an isolated
 CC beta1,3-N-acetylgalactosaminyltransferase (beta1,3GNT) or an active
 CC fragment, where beta1,3GNT directs expression of a MECA-79 antigen. The
 CC invention also provides a method of treating or preventing an
 CC L-selectin-mediated condition by reducing the expression or activity of a
 CC beta1,3GNT that directs expression of a MECA-79 antigen. This can be done
 CC by administering to the subject an oligosaccharide L-selectin antagonist
 CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
 CC administering antibody material that specifically binds beta1,3GNT,
 CC and/or a beta1,3GNT antisense nucleic acid molecule. L-selectin
 CC sulfotransferase-2 (LSST-2) also directs MECA-79 antigen expression.
 CC Alternatively, the expression or activity of LSST-2 or its active
 CC fragment can be reduced in combination with reducing the expression or
 CC activity of beta1,3GNT. The method is useful for treating L-selectin
 CC mediated conditions such as Crohn's disease and ulcerative colitis,
 CC inflammatory disorders of the skin such as allergic contact dermatitis,
 CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
 CC sequence represents human LSST-2.
 CC
 SQ Sequence 380 AA;
 Query Match 73.4%; Score 1498.5; DB 23; Length 380;
 Best Local Similarity 72.1%; Pred. No. 1e-150;
 Matches 277; Conservative 48; Mismatches 51; Indels 3; Gaps 2;
 QY 8 RLMLFSGQYIVVALFTIMSVHR--HLSQREESRRPVHVLVLSWRSGSSFPVGLFQGH 65
 DB 2 KLLFLVSGMAIILALFFHMYSHNLSLSMAQPER-MHVLVLSWRSGSSFPVGLFQGH 60
 QY 66 DVFYLMPEAHVMTFTSSSTAMKILMAVRLSLSVFLCDMSVPDAYNPGPKROSSLFPW 125

Db 61 DVFYLMERAMHVMWTFKOSTAMMLHMAVRLIRAVFLCDMSVFDAYMEGPRRQSSLCQM 120
 QY 126 EOSRALCSAPVCDFFPAHEISSPKCKILCGQPFDMVEKACRSHGFVYLKEVRELSIOA 185
 Db 121 EMSRALCSAPACDIIIPDEIIPRACHRLCSCQPFVEVEKACRSYSHVYLKREVENLOS 180
 QY 186 LYPILTDPSLNLHVHLVDRPRAVRSRREHTTIELMWSHVLGCHLETIKEEDOPYAM 245
 Db 181 LYPILKDPSSLNLHVHLVDRPRAVRSRREHTTIELMWSHVLGCHLETIKEEDOPYAM 240
 QY 246 KRICSOVDYKAIQTLPEALQOYTLFREDVLAAPLAQTRLKFKVGLDFLPHLOPMV 305
 Db 241 QVICSOLEIKYTIQSLPKALQERYLVRVEDLAPAPVAQTSRMTEFVGLEFLPHLOPMV 300
 QY 306 YNVRGKGMOGHAFHTNARNALNVSOAMRMSLPYERKVSQLODACEGEMDLGLGYRSGQ 365
 Db 301 HNITRGKMGDAFHTNARNALNVSOAMRMSLPYERKVSRLQACGDAMNLGVRHVRSGQ 360
 QY 366 EOGNLSLDLSSSHILGOV 384
 Db 361 EQRNLLDLSTWTVPEQI 379

RESULT 5

AAV79219
 ID AAV79219 standard; Protein; 386 AA.

XX AAV79219;

DT 19-JUN-2000 (first entry)

XX Human transferase TRNSFS-11.

XX Transferrase; TRNSFS-11; human; antitumor; cell proliferation;
 KW gastrointestinal disorder; developmental disorder;
 KW genetic disorder; neurological disorder; reproductive disorder;
 KW smooth muscle disorder; immunological disorder; inflammation;
 KW diagnosis; therapy; N-acetylglucosamine 6-O-sulfotransferase.

XX Homo sapiens;

XX Key Location/Qualifiers

FT Modified-site 121 /note= "potential O-phosphorylation"

FT Modified-site 107 /note= "potential O-phosphorylation"

FT Modified-site 217 /note= "potential O-phosphorylation"

FT Modified-site 252 /note= "potential O-phosphorylation"

FT Modified-site 364 /note= "potential O-phosphorylation"

FT Modified-site 380 /note= "potential O-phosphorylation"

FT Modified-site 35 /note= "potential O-phosphorylation"

FT Modified-site 50 /note= "potential O-phosphorylation"

FT Modified-site 81 /note= "potential O-phosphorylation"

FT Modified-site 287 /note= "potential O-phosphorylation"

FT Modified-site 243 /note= "potential O-phosphorylation"

FT Modified-site 30 /note= "potential O-phosphorylation"

FT Modified-site 308 /note= "potential N-glycosylation"

FT Modified-site 329 /note= "potential N-glycosylation"

FT Domain /note= "potential N-glycosylation"

FT /note= "transmembrane domain"

PN WO200014251-A2.
 XX 16-MAR-2000.
 XX 09-SEP-1999; 99WO-US20989.
 XX 10-SEP-1998; 98US-0150657.
 PR 04-NOV-1998; 98US-0186779.
 PR 11-MAY-1999; 99US-0133642.
 XX (INCYTE PHARM INC.
 PA Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H,
 PI Hillman JL, Azimzal Y;
 DR WPI: 2000-256996/22.
 DR N-PSDB; AA294211.
 XX Human transferase proteins useful for preventing, diagnosing and
 PT treating cancers and developmental, gastrointestinal, genetic,
 PT immunological, neurological, reproductive and smooth muscle disorders -
 PS Claim 1; Page 90-91; 113pp; English.

CC The present sequence is that of human transferase TRNSFS-11, 1 of
 CC 15 claimed novel human transferase proteins of the invention (see
 CC AAV79209-23). The sequence was deduced from a cDNA clone (see
 CC AA294211) isolated from a gallbladder library. It shows homology to
 CC mouse N-acetylglucosamine 6-O-sulfotransferase. TRNSFS-11 is
 CC expressed in dermatologic and gastrointestinal tissues,
 CC especially those associated with inflammation and cell
 CC proliferation. The new human transferases and polynucleotides can
 CC be used in the diagnosis, prevention and treatment of cancer,
 CC developmental disorders, gastrointestinal disorders, genetic
 CC disorders, immunological disorders, neurological disorders,
 CC reproductive disorders, and smooth muscle disorders. The
 CC polypeptides can also be used to raise antibodies, and to screen
 CC for agonists and antagonists of transferase activity.

XX SQ Sequence 386 AA;

Query Match 70.9%; Score 1448.5; DB 21; Length 386;

Best Local Similarity 71.1%; Pred. No. 2,2e-145; Mismatches 59; Indels 5; Gaps 4;

Matches 275; Conservative 48; Mismatches 59; Indels 5; Gaps 4;

QY 1 MMLKGRLLMFLGSOVIVALLFTHSVHR--HLSQRESRPVPVYLVSSRRSSSEVG 58

Db 1 MLLPKMKLLFLVQMAITLFFHYSNINISLSMKAPER-MHVLVSSWRSSSEVG 59

QY 59 QLFQGHDPVFLMEPRAMHVMWTFSTANKLMAVRLIRAVFLCDMSVFDAYMNPGRK 118

Db 60 QLFQGHDPVFLMEPRAMHVMWTFKOSTAMMLHMAVRLIRAVFLCDMSVFDAYMEGPR 119

QY 119 GSSLFOWEORALCSAPVCDFFPAHEISSPKCKILCG-QOPDMVEKACRSHGFVYLKE 177

Db 120 GSSLFOWEORALCSAPACDIIIPDE-SSPGITASSCAVNSLKLLEKACRSYSHVYLKE 178

QY 178 VRELSQALYPLTDPSSLNLHVHLVDRPRAVRSRREHTTIELMWSHVLGCHLETIKE 237

Db 179 VREFNLQSLYPLKDPSSLNLHVHLVDRPRAVRSRREHTTIELMWSHVLGCHLETIKE 238

QY 238 EDQPYTKAKITCSQVDYKAIQTLPEALQOYTLFREDVLAAPLAQTRLKFKVGLDE 297

Db 239 EDQPYTKAKITCSQVDYKAIQTLPEALQOYTLFREDVLAAPLAQTRLKFKVGLDE 298

QY 298 LPHLOTWYNNVRGKGMOGHAFHTNARNALNVSOAMRMSLPYERKVSQLODACEGEMDLGL 357

Db 299 LPHLOTWYNNVRGKGMOGHAFHTNARNALNVSOAMRMSLPYERKVSRLQACGDAMNLGL 358

QY 358 YLQVRSQOEGNLSLDLSSSHILGOV 384

Db 359 YLQVRSQOEGNLSLDLSTWTVPEQI 385

CC and detecting susceptibility to MCD. (I) is located to chromosome 16q22,
 CC and has ophthalmological activity. (I) can be used to treat or prevent
 CC macular dystrophy types I or II. (I) makes possible treatment
 CC of MCD without requiring keratinoplasty or keratectomy. The present
 CC sequence represents mouse intestinal N-acetylglucosamine-6-
 CC sulfotransferase, which is given in comparison with (I) in the
 CC exemplification of the present invention.

XX Sequence 418 AA;

Query Match 49.0%; Score 1001; DB 23; Length 418;

Best Local Similarity 54.5%; Pred. No. 1,2e-97;
 Matches 207; Conservative 45; Mismatches 118; Indels 10; Gaps 4;

QY 4 LKKGRLMFLGSGQVYVALFIHMSVHRHLSQRESRNP-----VHVLVLSWRSGSSSFV 57
 DB 21 LARGMRPLRFSSSTVMSLTMVGTGLVFLVSQVSSPAGLGERHVLVLSWRSGSSSFV 80
 DB 58 GOLFQHPDVFYLMERPAHVMWTFSTPAKMLHMAVRDLRSVFLCDMSVFDAYMNPGR 117
 DB 81 GOLFQHPDVFYLMERPAHVMWTFSTPAKMLHMAVRDLRSVFLCDMSVFDAYMNPGR 139
 QY 118 KOSLFEWESRALCSAPVCDFFPAHEISSPKHCKILCGQOPEDMVEKACRSHGFVYLKE 177
 DB 140 NISLDFQMAVSRAICSPVCEAFARGNISSEVCKPLCATRPFGLAQECSSYSHVYLKE 199
 QY 178 VFRLSLQALYPLLTDPISLTHVHLVRDPAVRSREHTTIELMVDSHVLCQHLETTKE 237
 DB 200 VAFRLQVLYPLSLPALNLRIVHLVROBRAVLRREGQAKALANDNGVLCGN-GTWE 258
 QY 238 EDQPYAKTIKCSQVDYKAIQTLPEALQORLYLFRLEDVLRAPLAQTRLYKVGDFE 297
 DB 259 ADPLRLVNEVCYSHRVAIEALHMKPPFLQDRYLRVREYEDLARDELTVIRELYAFTGICL 318
 QY 298 LPHLTQWYNTVRGKGMG--OHAFTNARNALNVSQAMRWSLPEYKVSQLODACEAMDL 355
 DB 319 TPQLDTWHTNTHGSGPCARRAEFTKTSRDALVSQAMRHTLPFAKIRVQELCGALQL 378
 QY 356 LGYLQVRSQOEGNLSLDL 375
 DB 379 LGYRSVHSELQORDSLDL 398

RESULT 8
 AAY72638

AAY72638 standard; Protein; 395 AA.

AAY72638;

02-MAY-2001 (first entry)

Mouse glycosyl sulfotransferase-4 (GST-4).
 Mouse; glycosyl sulfotransferase-4; GST-4; immunosuppressive;
 MM systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 MM polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 MM glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailitis;
 MM Hashimoto's disease; Grave's disease; hypoparathyroidism; anemia;
 MM demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 MM myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 MM asthma; hypersensitivity; rheumatic fever; tissue rejection;
 MM chromosome 8E1.

XX Mus musculus.

XX WO200106015-A1.

XX 25-JAN-2001.

XX 19-JUL-2000; 2000WO-US19741.

XX 20-JUL-1999; 99US-0144694.

PR 13-JUL-2000; 2000US-0593828.

XX (REGC) UNIV CALIFORNIA.

XX Rosen SD, Lee JK, Hemmerich S;

XX WPI; 2001-138471/14.

XX N-PSDB; AAD02696.

PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 XX diagnostic and therapeutic agent screening applications

PS Claim 3; Fig 2; 128pp; English.

XX The present sequence is mouse glycosyl sulfotransferase-4 (GST-4). GST-4
 CC gene is found on chromosome 8E1.
 CC GST is a type 2 membrane protein useful for inhibiting a binding event
 CC between a selectin and a selectin ligand, which comprises contacting the
 CC selectin with a non-sulphated selectin ligand, GST and a small molecular
 CC agent that inhibits the sulphation activity of GST. GST is also useful
 CC in inhibiting a selectin mediated binding event. GST is useful in gene
 CC therapy to treat disorders such as acute or chronic inflammation,
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adrenailitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation.

XX Sequence 395 AA;

Query Match 48.2%; Score 984; DB 22; Length 395;

Best Local Similarity 54.9%; Pred. No. 7.3e-96;
 Matches 207; Conservative 44; Mismatches 114; Indels 12; Gaps 6;

QY 8 RLMLFLGSGQVYVALFIHMSVHRHLSQRESRNP-----VHVLVLSWRSGSSSFVQOLF 61
 DB 2 RLPRF-SSTVMSLTMVGTGLVFLVSQVSSPAGLGERHVLVLSWRSGSSSFVQOLF 60
 QY 62 GHPDVFYLMERPAHVMWTFSTPAKMLHMAVRDLRSVFLCDMSVFDAYMNPGRKSS 121
 DB 61 SQHPDVFYLMERPAHVMWTFSTPAKMLHMAVRDLRSVFLCDMSVFDAYMNPGRKSS 119
 QY 122 LFEWESRALCSAPVCDFFPAHEISSPKHCKILCGQOPEDMVEKACRSHGFVYLKEVFTL 181
 DB 120 LFEWAVSRALCSPPVCEAFARGNISSEVCKPLCATRPFGLAQECSSYSHVYLKEVFTL 179
 QY 182 SLQALYPLLTDPISLTHVHLVRDPAVRSREHTTIELMVDSHVLCQHLETTKEEDP 241
 DB 180 NLQVLYPLSLPALNLRIVHLVROBRAVLRREGQAKALANDNGVLCGN-GTWEADPR 238
 QY 242 YRAMKTIKCSQVDYKAIQTLPEALQORLYLFRLEDVLRAPLAQTRLYKVGDFLPH 300
 DB 239 LRVNEVCYSHRVAIEALHMKPPFLQDRYLRVREYEDLARDELTVIRELYAFTGICL 298
 QY 301 LQTVYNTVRGKGMG--OHAFTNARNALNVSQAMRWSLPEYKVSQLODACEAMDL 358
 DB 299 LQTVYNTVRGKGMG--OHAFTNARNALNVSQAMRWSLPEYKVSQLODACEAMDL 358
 QY 359 LGYRSVHSELQORDSLDL 375
 DB 359 RSVHSELQORDSLDL 375

RESULT 9
 AAU11275

AAU11275 standard; Protein; 395 AA.

XX AAU11275;

Db	120	LFPMWNRALCSPPVEAARAGNISSEYCKRLCTRFPGIAQENCSYSHVVLKEVRF	179		
Qy	182	SLQALYPLLTDPDSLNLHVHVLVDRPRAVFRSREHTTEILMWDSHIVLGQHLETIKEEDP	2411		
Db	180	NLQVPLLTSDPALNRIYHVLVDRPRAVLRSPRECAKALARDNGIVLGTN-GTWVEADPR	238		
Qy	242	YVAMKIKCSQVDYIK-AIQITPEALQQRYLEFLREDELYVAPLQOTRLRYKFGVLDLPH	3000		
Db	239	LRVVNEVCSHRIAIAEALHMKPPPLQDQRYRLVREDELYAPLQVIRELYAFVGLGLTPQ	298		
Qy	301	LQTMVNVYRGKMGK--QHAFTNARNALNVSQARMSLPYKESQLQDADGCEAMDLLGY	358		
Db	299	LQTMVNIINIRGSPGARRAEKFTYTRDALSVSQARHRLPRAKIRVQELCGALQLLGY	358		
Qy	359	LYVRSQOEQGNLSLDLL	375		
Db	359	RSVHSELEQRDLSDLL	375		
RESULT 10					
ABB81555	standard; Protein; 395 AA.				
XX	ABB81555				
AC	ABB81555;				
XX					
DT	05-SEP-2002	(first entry)			
XX					
DE	Consensus N-acetylglucosamine-6-sulfotransferase SEQ ID NO:3.				
XX					
XX	Human: N-acetylglucosamine-6-sulfotransferase; enzyme: G1GNAC6ST;				
KM	Connell: sulfation; keratan sulfate; macular corneal dystrophy; MCD;				
KM	ophthalmological.				
XX					
XX	Homo sapiens.				
OS	Mus musculus.				
OS	Synthetic.				
XX					
EH	Key	Location/Qualifiers			
FT	Misc-difference 10	/label= Ala, Thr, Val			
FT	Misc-difference 13	/label= Ala, Val, Ser			
FT	Misc-difference 20	/label= Phe, Cys, Gly			
FT	Misc-difference 39	/label= Ala, Asp, Glu			
FT	Misc-difference 96	/label= Val, Met, Ile			
FT	Misc-difference 142	/label= Ala, Thr, Asn			
FT	Misc-difference 147	/label= Ala, Asp, Glu			
FT	Misc-difference 159	/label= Thr, Ser, Gly			
FT	Misc-difference 238	/label= Gly, His, Arg			
FT	Misc-difference 294	/label= Ser, Thr, Gly			
FT	Misc-difference 371	/label= Ala, Thr, Ser			
FT	Misc-difference 380	/label= Leu, Pro, Met			
FT	Misc-difference 382	/label= Gly, His, Ser			
FT	Misc-difference 384	/label= Thr, Ser, Lys			
FT	Misc-difference 390	/label= Ala, Glu			
FT	Misc-difference 391	/label= Ser, Lys			
FT	Misc-difference 392	/label= His, Gln			
FT	Misc-difference 394	/label= His, Gln			

QY	111	YANPGRROSSLTFQWEOESRALCSAPVCDPEFPAHEISSPKHCKLGGQDFDWEKACRSK	170
Db	110	YL-PKRRNLSDLEFQWAVSRALCSPPACSAFPPGAGISSSEAVCKPLCKROSFLLARECRST	168
QY	171	GFVVLKEVREYLSQALVPLLDPSLNLAHVVLVRDPAVERSRSEHTTTEIAMDVSHVILGO	230
Db	169	SHVVLKEVREYFNLQVLYPLDSDPALNLRIVHVLVRDPAVALRSREQTAKALARNGLVGT	228
QY	231	HLFETKEEDOPYAKKTIKCSQVDVLYKAIQTL-PEALDQRYLFLVEEDLYVAPLQOTTR	288
Db	229	N-GTWVEADPGLRAYRVEYCRSHVRYLAEAL-ATLKPPFLGRRIRLVFEDLAREPLAEYIRA	286
QY	289	LYKEVGLDPLPLHLQWTWYVNTVRGKGMG--QHAFTNARNALNVSQAMRSLSPREKYSOLO	346
Db	287	LYAEVGLSLTLPOLAEAMINHITHGSGPGAREAREAFKTTSSRALNVASQAMRHRLPFAKIRVQ	346
QY	347	DACGEAMDLTGILQVRSQOQGNLSDLL	375
Db	347	ELCAGALQLLGIRPYISDEQNRNLLDLY	375

XX	Sequence	390 AA:
Query Match	46.9%; Score 958.5; DB 22; Length 390;	
Best Local Similarity	55.6%; Pred. No. 3.8e-93;	
Matches 193; Conservative	42; Mismatches 87; Indels 25; Gaps 6	
QY	42 VHVILVSSWRSGSSFEVGLQFGHPDVFYLMEMBAHVMMFTTSSTANKLHMAVRDLLRSVF	101
DB	42 VHVILVSSWRSGSSFEVGLQFGHPDVFYLMEMBAHVMMFTTSSTANKLHMAVRDLLRSVF	101
QY	102 LCDMSVFPAYMNPGRKSSLEFQWQSRALCSAPVCDFEPFAHIESSPKKCKLLCGOQPD	161
DB	102 LCDMDVFAYM-POSRLNSAFFNMTSRALCSPPACSAFPGTISQODVCKTLCTQOPES	160
QY	162 MYEKARSGHGEVLTKEVRFSLQALYPLTIDPSLNVHVVLYVRDPRAVRSREHTTELM	221
DB	161 LARECRKRYSHVLTKEVRFNLIQVLYPLISDPAHLNRIYHLYVRDPRAVLRSRPAAGPILA	220
QY	222 VDSHVIIVG-----QHLETIKEEDDPYAMKIIICKSOVDIVKAIQTL--PEALQOORY	270
DB	221 RDNGVILGTNGKMWVEADPHRLRIRE-----VCRSHVRIAEK-ATLKPPPLLRGR	269
QY	271 LFLRYEDLVRAPLAOTRLRYKVGIDFLPHLOTWYVNTVRGKMGQ--HAFHTNARNALN	328
DB	270 RLVRREDLAREPELAEIRALYAFVTGLTLPQLEAMWIMINITHSGIGKPIEAFHTSSNARN	329
QY	329 VSOAMRMSLPYEKYSQLODAGCEAMDLLGVLQVRSQQECSNLSIDL	375
DB	330 VSOAMRHALPFTKILRVOECGALQLLGRVYSADQQRDLTLDDV	376

CC and detecting susceptibility to MCD. (I) is located to chromosome 16p22
CC and has ophthalmological activity. (I) can be used to treat or prevent
CC macular corneal dystrophy types I or II. (I) makes possible treatment
CC of MCD without requiring keratoplasty or keratotomy. The present
CC sequence represents human Intrastrinal N-acetylglucosamine-6-
CC sulfotransferase, which is given in comparison with (I) in the
CC exemplification of the present invention.

Sequence	390 AA;
Query Match	46.9%; Score 958.5; DB 23; Length 390;
Best Local Similarity	55.6%; Pred. No. 3,8e+93;
Matches 193; Conservative	42; Mismatches 87; Indels 25; Gaps 6

Search completed: August 9, 2003, 16:42:40
Job time : 40.6021 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:40:34 ; Search time 15.0388 Seconds
(without alignments)
2481.152 Million cell updates/sec

Title: US-09-645-078-4

Sequence: 1 MMLKKRLLMFLGSGVIV.....NLSDLLSSHLIGQVREG 388

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616882 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

1: PIR-76:*
2: PIR1:*
3: PIR2:*
4: PIR3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	593	29.0	486	2 JC7351	N-acetylglucosamin
2	591.5	28.1	484	2 JC7350	N-acetylglucosamin
3	574.5	28.1	484	2 JE0261	N-acetylglucosamin
4	496.5	24.3	458	2 A57397	chondroitin 6-sulf
5	134.5	6.6	388	2 G70729	hypothetical prote
6	124	6.1	307	2 E95934	probable enzyme, C
7	101.5	5.0	4273	2 C69679	polyketide synthas
8	98.5	4.8	3187	2 JC5837	364k Golgi complex
9	97.5	4.8	291	2 F73826	Me240 homolog F10-
10	94.5	4.6	513	2 F90866	transcription regu
11	93	4.6	346	2 E86319	probable flavonol
12	93	4.6	726	2 T35865	probable cyclase -
13	92.5	4.5	292	2 AD0362	probable lysr-fami
14	92.5	4.5	513	1 REGCAY	transcription regu
15	92.5	4.5	1495	2 S61023	hypothetical prote
16	92	4.5	812	2 D84964	transcription-repa
17	91.5	4.5	338	2 A96769	protein flavonol s
18	91.5	4.5	513	2 C85752	hypothetical prote
19	91.5	4.5	1169	2 F97872	transcriptorepair
20	91	4.5	2354	2 T13288	mei-41 protein - f
21	90.5	4.4	1169	2 F95000	transcription-repa
22	89	4.4	361	2 T41176	hypothetical prote
23	89	4.4	910	2 E89918	2-oxoglutarate deh
24	88	4.3	469	2 D84857	hypothetical prote
25	88	4.3	905	2 B71562	probable salt-indu
26	87	4.3	822	2 T01622	probable salt-indu
27	86.5	4.2	467	2 F83651	hypothetical prote
28	86.5	4.2	664	2 T48258	kinesin-like prote
29	86.5	4.2	892	2 T27005	hypothetical prote

30	86.5	4.2	1116	2 H97970	type I site-specif
31	85.5	4.2	974	2 A44484	transcription regu
32	85.5	4.2	1091	2 B95103	hypothetical prote
33	85	4.2	344	2 C96599	protein F14J16.8 f
34	85	4.2	844	2 C37794	aminopeptidase Y9c
35	84	4.1	995	2 A84014	SNF2 helicase BH29
36	84	4.1	1293	2 T27886	hypothetical prote
37	84	4.1	1813	2 T19295	hypothetical prote
38	83.5	4.1	424	2 A71476	probable protopor
39	83.5	4.1	623	2 G81420	hypothetical prote
40	83.5	4.1	960	1 OX1X2	phosphoenolpyruvat
41	83	4.1	398	2 C86749	hypothetical prote
42	83	4.1	513	2 AE0659	transcription regu
43	83	4.1	655	2 G87306	threonyl-tRNA synt
44	83	4.1	776	2 T20896	hypothetical prote
45	83	4.1	1568	2 T41013	hypothetical prote

ALIGNMENTS

RESULT 1

JC7351 N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human

C:Species: Homo sapiens (man)

C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000

C:Accession: JC7351

R:Uchimura, K.; Fasakhany, F.; Kadomatsu, K.; Matsukawa, T.; Yamakawa, T.; Kurosawa, Blochem. Biophys. Res. Commun. 274, 291-296, 2000

A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of

A:Reference number: JC7350; MUID:20374462; PMID:10913333

A:Accession: JC7351

A:Molecule type: mRNA

A:Residues: 1-486 <UCH>

A:Cross-references: DDBJ:AB040711

C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosyl

sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand

C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 29.08; Score 593; DB 2; Length 486;

Best Local Similarity 37.78; Pred. No. 4.6e-44;

Matches 146; Conservative 51; Mismatches 136; Indels 54; Gaps 10;

QY	38	SRRPVHVLVSSWSSGSPVGLFGHDPVYFLMEPAHVMWMTSSPAKMLHVAVDL	97
DB	98	SREKQHTVHAIVMTGSSFLGELNQHDPVFLTEPMHMLQALYPGDAESLQALKDL	157
QY	98	RSVFLCDMSYFDAYMNG-PRKQ-----SSLFQWESRALCSAPVCDFFPAHE---	144
DB	158	RSLEKCDPFSVLRLVAPGDPAPARAPDTANLTALFMFRNKVICSPPLCGAPARA	217
QY	145	-ISSPKCKLILCGQPPDMWEKACRSHGFFVLAKVRLSIALYPLLTDSLNHVVHL	203
DB	218	GLVEDTACERSCPVAIRALEAECKRYVVIKQVRLDLGLVPLRLDPLNLKVOLF	277
QY	204	RDPRAVRSREHTEILMDS-----HYVGQH-----LETKEE	238
DB	278	RDPRAVNSRLKSKQGLRESIOVLRKRGDRHRYVLHGVARGGSRALPAAPRA	337
QY	239	DOPYR---AKIICKSQVDIVKATQTLPEALQRYLFLREEDLVRAPLAQTTLRYK	295
DB	338	D--FELGALVEICEALRLDLFARGAPAMLRRLRYLRLREDDVLRQRAQLRLRS	395
QY	296	DPLPLOTWYVNTYTRGGMG-QIAFHNARNALNVSAWKRSLEFVYSQADAGC	354
DB	396	RALNALDPAFLNMTRGAYGADRPFLSARDAEVAHMBRLSREQVQVEACAPAMR	455
QY	355	ILGAYVRSOQ-----OGNLSLDL	374
DB	456	ILAY--FRSGEEDGADQPRECEPLEM	480

RESULT 2

N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse

C:Species: Mus musculus (house mouse)
C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000
A:Accession: J07350

R:Uchiyama, K.; Fasaki, H., F.; Kadomatsu, K.; Matsukawa, T.; Yamakawa, T.; Kurosawa, N.
Biochem. Biophys. Res. Commun. 274, 291-296, 2000

A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A:Reference number: J07350; MUID:20374462; PMID:10913333
A:Accession: J07350

A:Molecule type: mRNA
A:Residues: 1-484 <UCH>

A:Cross-references: DDBJ:AB040710
C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 28.1%; Score 591.5; DB 2; Length 484;
Best Local Similarity 38.1%; Pred. No. 6,2e-44;
Matches 136; Conservative 49; Mismatches 111; Indels 41; Gaps 6;

OY 43 HVLVSSRRGSSSVGGQFGCHDPVFYLMEPAHVMWTFSTAMKLMAYRDLRSVFL 102
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 101 HIYHAATWRGSSSGELFNGHPDVFYLEPBMHLMOALYPGDASELQGLARLDRLSLFR 160
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 103 CDMSVPFAVNMWG-PKRO-----SLEFMQDSALCAPYCDEFPNHE-----ISSP 148
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 161 CDSEVLRILYAPGPGGEAPDSANLTTRAMEFRMTNNVCSPPLCPAARRADVDGED 220
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 149 KHCKILCGQGFDDVEAKERSHGFEVFLKEVAFSLQALYLPRTPSINTLHVHLVVRPPA 208
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 221 KACSTCPBPVALALEACRKRPVVYIKDVALLDGLVPLRLRGILTAKVQLFRPRRA 280
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 209 VFSREHTTIELMDSHYLV----GHLEFI-----KEEDOPY 242
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 281 VNHSRLKSROGLRESIOVLTQRGDHFHFVLLAHGVDAARPQAALBSAPPADFELT 340
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 243 YAMLIICKSOVDIYKALOTLPBALQORYLFERYEDLVAPLAOTRIKYKFGIDFLPHIQ 302
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 341 SALEVICBAMLRDLFTFRGAPAMLRRLIRLYTDVLWQPAAQLRRLIRSGELTLAALD 400
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 303 TWNVNTVRGKGMG-QHAFTNRARNALNVSQAMRSLEYEKVSQQLDAGCEAMDILIGY 358
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 401 AFAENMTGRSAVGADRFPHLSARDAREAVHWRRRLSQEQRYQVETACAPMRRLAY 457
| : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 3

J0E0261
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human
C:Species: Homo sapiens (man)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 01-Mar-2002
A:Accession: J0E0261

R:Uchiyama, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Fan, Q.W.; Mitsunaka, J.
J. Biochem. 124, 670-678, 1998

A:Title: Human N-Acetylglucosamine-6-O-sulfotransferase involved in the biosynthesis of
A:Reference number: J0E0261; MUID:98391845; PMID:9722682
A:Accession: J0E0261

A:Molecule type: mRNA
A:Residues: 1-484 <UCH>

A:Cross-references: DDBJ:AB014679
C:Comment: This protein catalyzes the transfer of sulfur of sulfate from 3'-phosphoadenine C:Superfamily: Chondroitin 6-sulfotransferase
C:Keywords: sulfotransferase

Query Match 28.1%; Score 574.5; DB 2; Length 484;
Best Local Similarity 35.0%; Pred. No. 1.9e-42;
Matches 130; Conservative 65; Mismatches 121; Indels 55; Gaps 9;

OY 46 VLSSMRGSSSVGGQFGCHDPVFYLMEPAHVMWTFSTAMKLMAYRDLRSVFLCDM 105
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 123 VFTWRGSSSFEGELFNONEVEFFLYEPVWHWOKLYPGDAVSLQGAARDMLSLVRCDL 182

[illegible]

Db 449 L 449

RESULT 5

G70729 hypothetical protein RV2267c - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: G70729

R:Colet, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

R:Connot, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.

R:Rajendram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

R:Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:96295987; PMID:9634230

A:Accession: G70729

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-388 <COL>

A:Cross-references: GB:277163; GB:AL123456; NID:93261610; PIDN:CAB00968.1; PID:e255072;

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: RV2267c

Query Match

Best Local Similarity 21.8%; Score 134.5; DB 2; Length 388;

Matches 77; Conservative 59; Mismatches 116; Indels 101; Gaps 20;

QY 44 VLVLSWRSGSSFFVGOLF--GQH-----PDVFYLMEPAMHV-WMTFTSSTWAKL 89

Db 83 IFYGHWRGTGTHLHLLVVDHRTGPTGECLEAPHHFLTE--WPAVYEFELVSK----- 136

QY 90 HMAVRDLRSVFLCDMSVFADYANPGRKQSLQWQESRALCSPVCFPAHEISSPK 149

Db 137 HRADNN-----DLSLH-----HPQDEFVW-CMOGLPSYLTIAFNNR---PPQ 177

QY 150 HCKLLCGQPFDPWEKAKRSHGFVLEKVEFLSLQALY-----PLTDPSLNHV--- 199

Db 178 Y-----EYKLDQVAPRE--LEIKWRTLEFRVQVYFRKRVYILKNPHSRIVKL 229

QY 200 -----VHLRDPRAVFRSREHTTELWDSHIVLGOHLETKEDQPYAMKIKKS 251

Db 230 EVFQAKFIHVRDPPVYVPS---TILHKAFLYRIGLQPTFDGDK-----VYST 279

QY 252 QVDVKAIGTLPELQ-QRYLFLEEDLVAPLAQTRIKYFGLD---FLPHLQWVY 306

Db 280 YVDLKRRLDEGRLEVDPTREYELRELDIGPQGLRXLXOHLGDFECCYLPRLQY-- 337

QY 307 NVTRGKMGQHA-FHTNARNALNVSAQMRSLPEKYSQLODAGCEAMDILGY 358

Db 338 -----LADHADYKINS-----YGLIVEQRAIVDEHGEIIDRIGY 372

RESULT 6

E95934 probable enzyme, C-terminal domain similar to sulfoltransferase protein [imported] - Sinc

C:Species: Sincrobacterium meliloti

C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: E95934

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernat

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1.683-Kb PSYMB megaplasmid from the N2-fixing endo

A:Reference number: A95842; MUID:21396508; PMID:11481431

A:Accession: E95934

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-307 <KOR>

A:Cross-references: GB:AL591985; PIDN:CAC49141.1; PID:915140626; GSPDB:GN00167

A:Experimental source: strain 1021, megaplasmid pSYMB

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

Pela, D.; Chain, P.; Cowle, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau

hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yen,

A:Title: The composite genome of the legume symbiont Sincrobacterium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: Smb21237

A:Genome: plasmid

Query Match

Best Local Similarity 21.3%; Score 124; DB 2; Length 307;

Matches 79; Conservative 54; Mismatches 120; Indels 118; Gaps 19;

QY 40 RPVHVLSSW-RGSSFFVGOLFQHPDVF--YLMEPAMHV-----MTF 81

Db 4 QPVIAYIAGYGRSGSTILDIALGQHAHVAGETTSITRHVHRHNEACAGNAIRDCSF 63

QY 82 TSSP--AMK-----LHMAVRDLRSVFLCDMSVFADYANPGRKQ 119

Db 64 WSSVREWSGQDPGLMEYCALQOKREGLSMTR-LISGN-----GLGKQ 108

QY 120 SLPQWESRALCSPVCFPAHEISSPKHCKLLCGQPFDPWEKAKRSHGFVLEKVR 179

Db 109 FSLYTLHTRKLFSAHQSS-----GRQ---YVDSKLPG-----R 141

QY 180 FLSLQALYPLTDSNLHVVHLYRDPRAVFRSREHTTELWDSHIVLGOHLET---IK 236

Db 142 AMAVAOI-----GQIDRVYTHLVDRGQVAMS-----LLKGERAKSGIQ 182

QY 237 EEDQPYAMKIRK-SQVDYKAIOTLPEAL-QORYLFLEEDLVAPLAQTRIKYFVG 294

Db 183 KEIKPKSVFKRLKMSVNL--AVBYLSRKLGSEKVMKVRKEDASDPVAMQIGTFLE 240

QY 295 LDPLPHLQWVYVNTRGKMGQ-----HAETNARNALNVSAQMRSLPEKYSQLODA 348

Db 241 LD-LSQVTSLEN--GEAMGPHQVAGNRLRMANASIALNDETFRTMPARQGVSPQRL 296

QY 349 CGEAMDILGYL 359

Db 297 GGMMLRRYGYL 307

RESULT 7

C69679 polyketide synthase pksM - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 03-Nov-2000

C:Accession: C69679

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Erttington, J.; Fabel, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

lechi, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, I.

Koetter, P.; Konligstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A:Authors: Lander, U.J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

y, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl

A:Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpiltra, P.; Tognoni, A.; Tosato, V.; Uchlye

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Zanchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis

A:Reference number: A95380; MUID:96044033; PMID:9384377

A:Accession: C69679

A:Status: preliminary; nucleic acid sequence not shown; translation not shown.

A:Molecule type: DNA

A:Residues: 1-4273 <RUN>

A:Cross-references: GB:299113; GB:AL009126; NID:92634090; PIDN:CAB13603.1; PID:926341

A:Experimental source: strain 168

C:Genetics:

A:Gene: pksM
 C:Superfamily: Bacillus subtilis polyketide synthase pksM; 3-oxoacyl-[acyl-carrier-prote
 C:Keywords: carrier protein
 F:293-363/Domain: acyl carrier protein homology <ACAP>
 F:214-829/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OASI>
 F:2340-2732/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
 F:3145-3320/Domain: short-chain alcohol dehydrogenase homology <SAD>
 F:3350-3842/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
 F:4138-4208/Domain: acyl carrier protein homology <ACP>

Query Match 5.0%; Score 101.5; DB 2; Length 4273;
 Best Local Similarity 22.3%; Pred. No. 11;
 Matches 59; Conservative 44; Mismatches 99; Indels 63; Gaps 14;

133 LIGGQPPDWEKACRSHGFVLEKREPLSLQALYPLTDPSS---LNLHVHLYVLRDPA 208
 DB LIPGLAYIDLYQVQVEHGY-AYQELKLNLTIFVPLADESYDIALFTIH----- 74
 209 VFRSHHTIELMVSHVILGHTETKE-----EQPYAMKIIKQVDIVK 257
 DB --SEEEGTSWIIIDGQ---KQHGSLSDKROYETADMRKQDTAFESIDLNQKSTAD 129
 258 AIQTPLEALQORLYFLRYEDLVRAPIAQT-RLYK-----FVGLDFLPHLQTVYN 307
 DB RLINDEIYEQ---CRQELVHTGMKAEQIYEAKEGAVIDLAVGQDALHNSDAFLH 185
 308 VT--RGKMGQHAFTNARNALNVSQAMRSLPYEKVS---QLDAGCEAMDLLGYLYR 362
 DB PLIDGSGIGS-----SCLISDQTMVLPYEFSEFSERLQKGC--TANILS-SSVR 234
 363 SQOEGGNISDLSSSHLGOVFE 387
 DB 235 QKRELTYMTIEYFNSA---GQRVAE 256

RESULT 8

364K Golgi complex-associated protein - rat
 JCS837
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000
 C:Accession: JCS837
 R:Toki, C.; Fujiwara, T.; Sohma, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.
 Cell Struct. Funct. 22, 565-577, 1997
 A:Title: Identification and characterization of rat 364-kDa Golgi-associated protein reo
 A:Reference number: JCS837; MUID:98093490; PMID:9431462
 A:Accession: JCS837
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-3187 <TOK>
 A:Cross-references: DDBJ:D25343; NID:9516825; PIDN:BA05026.1; PID:9516826
 C:Comment: This protein plays a role in the formation and maintenance of the characteris
 C:Superfamily: giantin
 F:49-549,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper #status predict
 F:3165-3187/Domain: membrane anchor #status predicted <AAD>

Query Match 4.8%; Score 98.5; DB 2; Length 3187;
 Best Local Similarity 19.9%; Pred. No. 14;
 Matches 55; Conservative 54; Mismatches 93; Indels 75; Gaps 12;

149 KHCKLGGQPPDW---EKACRSHGFVLEKREPLSL---QALYPLTDPSSLNHYV 200
 DB 2286 QNCKNCRQLETDLVAASRLTTLRLHDEINVKQKIIISLSGKEEALIVAVIAE---LHQO 2341
 201 HL--VRDPAVFRSHHTIETL-----MVDSHIVLGOHLETKEDQPYAMKIIKQSQ 252
 DB HSKELKELENLISGEENLTLSEENKRAVEKTNLTALETIKKE-----SLEQKQ 2394
 2342 233 VD-IYKAQTLPE-----ALQQRYL--FLRYEDLVRAPIAQTRLKYFVGLDFLP 299
 DB 2395 LDFEFKSSSLQDDDRIVSDYRQLEERHLSVLEKDELIDQAAENNNKKEI----- 2448
 300 HLGTFVYVNTRGKGGQHAFTNARNA-----LNVSQAMRSLPYEKVS 343

DB 2449 -----RGLRGHMDINSENAKIDAEILQYRRDLNFTIKSQORQLLEAQIQ 2496
 OY 344 QLQDAGCEAMDLLGYLYVRSQOEOG-NISDLSSSH 379
 DB 2497 QNKELRNECVKLEGRKSGEAKQSLQNSLDALQGEN 2533

RESULT 9

573826
 MG240 homolog F10.orf291 - Mycoplasma pneumoniae (strain ATCC 29342)

C:Species: Mycoplasma pneumoniae
 A:Variety: ATCC 29342
 C>Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
 C:Accession: 573826
 R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
 Nucleic Acids Res. 24, 4420-4449, 1996
 A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumon
 A:Reference number: 573327; MUID:97105885; PMID:8948633

A:Accession: 573826
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-291 <HIM>

A:Cross-references: EMBL:AE000049; GB:U00089; NID:91674188; PIDN:AA96148.1; PID:9167
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
 C:Genetics:
 A:Genetic code: SGC3
 A:Superfamily: hypothetical protein ygeK

Query Match 4.8%; Score 97.5; DB 2; Length 291;
 Best Local Similarity 22.8%; Pred. No. 0 72;
 Matches 59; Conservative 43; Mismatches 94; Indels 63; Gaps 14;

OY 109 DAYNMPGRKQSSLEPQW---EQSRALCSAPVDFP--PAHEISSPKHCKLGGQPPDW 163
 DB 41 ELYPLIGSDKLSLAKWNHIEQLKICFRVCYERGRYPIDQLVOGFVRLKGCPLDIA 100
 OY 164 -EKACRSHGFVLEKREPLSLQALYPLTDPSSLNHYVLRDPAVFRSHHTIETLV 222
 DB 101 SSEMGSHNF---ROI-----PAKVLHYTH-----OH----- 124
 OY 223 DSHIVLGOHLETKEDQPYAMKIIKQSOVDIVAICPLPALQORYFLRYEDLVRA 281
 DB 125 --NITLKITLQTLDEPPRQHCRLV---GQLAKTLAVANKLDG-KTAYTAGVHDLAKQ 178
 OY 282 PLAQTRLYKFEVGLDFLPHLQTV--YVNTRGKMGQHAFTNARNALNVSQAMRSLPYE 340
 DB 179 POALEKLEKLVAGVNDYP---SMKVLIHSYAGVILKHWYGLN--NSAVSFALMNTVTPQ 233
 OY 341 KVSQQLQDAGCEAMDLLGYL 359
 DB 234 KMSQL-----DWIIVL 244

RESULT 10

F90866
 transcription regulator tyrr Ecs1902 [imported] - Escherichia coli (strain 0157:H7, s

C:Species: Escherichia coli
 C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Mar-2003
 C:Accession: F90866
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
 gatawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and 9
 A:Reference number: A59629; MUID:21156231; PMID:11258796
 A:Accession: F90866
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-513 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BA935325.1; PID:91361367; GSPDB:GN00154
 A:Experimental source: strain 0157:H7, substrain RMD 0509952
 C:Genetics:

A:Gene: Ecs1902
 C:Superfamily: transcriptional regulator of tyrr type with ACT, PAS, AAA, and FIS dom

Query Match 4.6%; Score 94.5; DB 2; Length 513;
 Best Local Similarity 19.4%; Pred. No. 2.8;
 Matches 84; Conservative 69; Mismatches 160; Indels 119; Gaps 20.

QY 5 KKGRLMFLGSOVIVVALFTHMSVHRLSQRESRRPVHVLSSWSSGFFVQLGQH 64
 DB 116 KLRLRNHTAQLINGNFL-----RMLESEPODSHNEHYI-----NGQNFLEI----- 161
 QY 65 PDVYTLMEPAMHVMHTTSSITAMKLMHVR--DLRSVFLCDMSVFDAVYMPGPKOSSL 122
 DB 162 TPVYLODENDQHV-----LTGAVMILRSTIRMGROLQVAAQDVSAFSQIVAVSP-KMHV 216
 QY 123 FOMEOSRALSAPVC-----DFF--PAHEIS--SPKHCILGQGOQFDMVEKACR 168
 DB 217 VEQAKLAMLASAPLLITGDTGTGDLFRAYACHQSPRASKRYLALNCSTIDEDVESELF 276
 QY 169 SH-----GFVYLKEVRLSLQALYPLTDPSSLNHVHVLVRDPRAVFRSRE 214
 DB 277 GHAEKKGFEEQANGSGSVLLDEIGEMS-----PRMQAKLRLRLND--GTFR--- 321
 QY 215 HTTELAVDSHIVGQHLETKEDQPYAMKTIICKSQVDYKAIQILPEALQRYFLR 274
 DB 322 -----RVGEDVHVDVAVICATQKNLVELVQ--KGMFREDLYXR 359
 QY 275 YEDLV--RAPLAQTRRLKFEVGLDFPLHLOTWVYNTVNGKMGQHAFTNARNALNVSQA 332
 DB 360 LNVLTMLPLRDCPQ-----DIMPLETEFVARFADQGVPRPKLAADINTVLT----- 408
 QY 333 WRMSLPYERVSQLODACEAMDLYLQVRSQ-----QEOGNSLIDLSS 377
 DB 409 -RYAMP-GMVFOLKNAIYRALQLDGYELRPDILLPDYDAVAVGDEAMEGSLDEITS 466
 QY 378 ---SHIIGOVER 386
 DB 467 RERSVLTQLXR 478

RESULT 11
 E86319
 Probable flavonol sulfotransferase [Imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: E86319
 R:Theologian, A.; Ecker, J.R.; Palm, C.J.; Federpiet, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huitz, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzalli,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11330712
 A:Accession: E86319
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-346 <STO>
 A:Cross-references: GB:AE005172; NID:99795597; PIDN:AAF98415.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1
 C:Superfamily: alcohol sulfotransferase

Query Match 4.6%; Score 93; DB 2; Length 346;
 Best Local Similarity 18.7%; Pred. No. 2.2;
 Matches 75; Conservative 58; Mismatches 149; Indels 120; Gaps 16;

QY 18 IVVALFTHMSVHRLSQRESRRPVHVLVS-----SMRSGSSFVQLFGQHPVYFLME 72
 DB 9 VVVSSESHELASSSPSEFEKQKHQYQELIATLPKDKWRKPDVPE--YGGHMMQLDPLE 66
 QY 73 PAMHVMTF-----TSSTAMKLMHVRDLRLSVFLCDMSVFDAVYMPGPKQ 119

DB 67 GLLAQKFKRPNDFVCSYPRKGTITWKL-----TFAINRSRFDVSTPFLKRN 119
 QY 120 SLSFOWEOSRALSAPVC-----FPAHEISSPRKHCILGQGOQFDMVEKACRSHGFVL 175
 DB 120 PHEF-----VPYIEIDFPFPPSVADVLDKDGNTLFSTHPIYDILLPEVSVSG----- 165
 QY 176 KEVFLSLQALYPLTDPSSLNHVHVLVRDPRAVFRSREHTITELMDSHVLQGHLETI 235
 DB 166 -----CKTYIYWRDKDF-----VSMWTFHAKERSQOQGPV 197
 QY 236 KEEDQPYAMKTIICKS-----QVDIV-----KAIQTLPEALQRYFLRYEDLVRAPIAQ 285
 DB 198 SIEE-----AFPKYCGSLASVGPYLDHVLGYKAKAQNAPDQ-----LFLKTYTMAADPLPY 249
 QY 286 TTRLYKTVGLDF-----LPHLOTWVYNTVNGKMGQHAFTNARNALNVS 330
 DB 250 VKRLAEFGYGFTEEBEGNVYERVKLSFET-LKMLEANKGEKR-----EDRPAYAN 304
 QY 331 QA-----WRMSLPYERVSQLODACEAMDLYLQVRS 363
 DB 305 SAYFRKGVGDWNTLTPEVARIIDGLMEERFKGTGFLSSKS 346

RESULT 12
 T35865
 Probable cyclase - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T35865
 R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajendram,
 submitted to the EMBL Data Library, April 1999
 A:Reference number: Z21591
 A:Accession: T35865
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-726 <SAU>
 A:Cross-references: EMBL:AL049727; PIDN:GAB41566.1; GSPDB:GN00070; SCODEB:SC9B1.20
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCODEB:SC9B1.20

Query Match 4.6%; Score 93; DB 2; Length 726;
 Best Local Similarity 19.8%; Pred. No. 6;
 Matches 79; Conservative 46; Mismatches 120; Indels 154; Gaps 19;

QY 36 EESRPVHVLV-----LSSWSSGSSFFVQLFGQHPVYFLMEPAMHVMHTTSSITAMK 88
 DB 291 EKGTLPLEVAAGVATKGLQDMQSGG-----HEW----- 319
 QY 89 LHMVRLRLRSVFLCDMSVFDAVYMPGPKOSSLSFOWEOSRALSAPVCDFPAHEISSP 148
 DB 320 HM-----RS-----SRYNKKGR--PLAQWQ-----ALTGP 343
 QY 149 KHCILGQGOQFDMVEKACRSHGFVLKEVR--FLSLQALYPLTDPSSLN--LHVHVL 202
 DB 344 GTSAADGALLADVAAGARARITYVPQKQVPSVIPDRMRYPELSPALDGARRHLSW 403
 QY 203 VRDPRAVFRSREHTITELMDSHVLQGHLETKEDQPYAMKTIICKSQVDYKAIQTL 262
 DB 404 CRE-----MGLISGVWDEDELDSCDPLCAAGID----- 433
 QY 263 PEALQOR-----YLFRLY--EDLVARPLAQTRRLKFEVGLDFLP----- 299
 DB 434 PDATQDDLDLASGLAFGTGDDYPLVYIGRRDLAARLT--TTRLSDCMFLDGEVPPP 492
 QY 300 -----HLQTWVYNTVNGKMGQHAFTNARNALNVSQARMSLPYERVSQLODACEA 352
 DB 493 GNAMERSLIDLVMTTA--GMPTEERRPLKKAQVDMTEAMTWELS-----NQIONRPPDP 545
 QY 353 MDLIGYLQVRSQEOGNSLIDLSSSH--ILGOVFRFG 388
 DB 546 VD--VLEMRRAFGSDLTGLICRAHGPAVPRVYVSG 581

RESULT 13

AD0362

probable lyxR-family transcription regulatory protein YPO2979 [Imported] - Yersinia pestis
 C:Species: Yersinia pestis
 C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C/Accession: AD0362
 R:Fairhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,

Nature 413, 523-527, 2001

A>Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AD0362

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-292 <YR>

A:Cross-references: GB:AL590842; PIDN:GAC92223.1; PID:g15980935; GSPDB:GN00175

A:Genetics:

A:Gene: YPO2979

Query Match

Best Local Similarity 4.5%; Score 92.5; DB 2; Length 292;
 Pred. Local Similarity 23.7%; Pred. No. 2;

Matches 50; Conservative 25; Mismatches 77; Indels 59; Gaps 10;

QY 184 QALYPLTPSLNHLVHVRDPRVFRSREHT-----IE-----LWDSHIVLG 229

DB 53 QARQPLVD-----HGRRLVAVHVOETLASERLTDLIRIAGOVERLTFVLSDIYQPA 107

QY 230 QHLETKEDDOPYAMKIIIC--KSGVDYKAIOTLPEA-----LQORFLFYREDLYVA 281

DB 108 HHENLIRREQRYPRVEFCVLADEEDVDILQ--LKRATIGVEMQER-----PPDIAT 162

QY 282 PLAQTRRLKYV---GIDFLPHLQTVYVNTVRGKMGQHAFTNARNALNSQAWMSL 337

DB 163 RLAAQTEMAVFLRGHPLRLPHLQNEQLATFR-----QLCLNTVYRTRDHNSQGLFWSA 217

QY 338 PYEKVSQLODAGCEMDLGLYQVRSQDQSG 368

DB 218 P-----SYLLEMAEOG 230

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

A:Accession: F64881
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA

A:Residues: 1-513 <BIAT>

A:Cross-references: GB:AE000230; GB:U00096; NID:g1787578; PIDN:AACT4405.1; PID:g17875

A:Experimental source: strain K-12, substrain MG1655

A:Genetics:

A:Gene: tyxR

A:Map position: 29 min

A:Complex: homodimer

C:Function:

A:Description: regulates expression of genes involved in aromatic amino acid biosynth

A>Note: autogenously regulated; activity depends on concentrations of aromatic amino

C:Superfamily: transcriptional regulator of tyxR type with ACT, PAS, AAA, and FIS dom

C:Keywords: DNA binding; nucleotide binding; P-loop; transcription regulation

F:206-421/Domain: RNA polymerase sigma factor interaction domain homology <SF1>

F:234-241/Region: nucleotide-binding motif A (P-loop) #status atypical

F:294-298/Region: nucleotide-binding motif B

Query Match

Best Local Similarity 4.5%; Score 92.5; DB 1; Length 513;
 Pred. Local Similarity 19.4%; Pred. No. 4.2;

Matches 84; Conservative 71; Mismatches 157; Indels 121; Gaps 21;

QY 5 KKGRLMFGSQVIVVALEFIMSVHRHLSQRESRPRVYVLYLSSMRGSSFPVGLFGH 64

DB 116 KIDRLRNHTAOLINGFNL-----RWLESEPDOSHNEHVY-----NGQNTLMEI----- 161

QY 65 PDVFLMEPRAMVMTFTSTAMKLMMAVR--DLRSVFLCDMSVFDAYMNPGRKOSL 122

DB 162 TPVYIADENDQHV-----LTGAVVMLRSTRMGRQLQNVAAQVSAFSGVANSF--KMKHV 216

QY 123 FQWESRALCSAPVC-----DPF--PAHEISSPK-----HCKLLCGQDFDVEAKC 167

DB 217 VEQAKLMLASAPLLITGTGTGKDLFAVACHQ--ASPRGKRYLALNCASIPEDAVESEL 275

QY 168 RSH-----GFVLEKVRFLSLALYPLDPSLNHLVHVRDPRVFRSR 213

DB 276 FGHAEGRKGFPEQANGSVLDEIGEMS-----PRQALFLIND--GTFR-- 321

QY 214 EHTTLEMDSHIVLQHLETKEDDOPYAMKIIICSGVDYKAIOTLPEALQGRHFL 273

DB 322 -----RVGEDHEVAVDVAVICATQKNVLELVQ--KMFREDLYX 358

QY 274 RYEDLV--RAPLAQTRRLKYVFGIDFLPHLQTVYVNTVRGKMGQHAFTNARNALNSQ 331

DB 359 RLNVLTLLNPLRLDCPQ-----DIMPLETVFVARFADQGVPRPKLAADLNTVLT--- 408

QY 332 AMRWSLPEYKVSQLODAGCEMDLGLYQVRSQ-----QEOGNTSLDLIS 376

DB 409 --RYAMP--GNVQRLNATYRALTOLDGYELRQDILLPDYDAATVAVGEDAMESSLDEIT 465

QY 377 S---SHILGOVFR 386

DB 466 SRFERSVLTOLYR 478

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:39:29 ; Search time 9.52455 Seconds
(without alignments)
1915.719 Million cell updates/sec

Title: US-09-645-078-4
Perfect score: 2042
Sequence: 1 MALLKGRILMFLGSGQIVV.....NLSDLLSSHIILGQVFRG 388

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	496.5	24.3	458	1 C6ST_CHICK	092179 gallus galli
2	134.5	6.6	388	1 YME7_MYCTU	050695 mycobacteri
3	101.5	5.0	4273	1 PKSM_BACSU	P40872 bacillus su
4	97.5	4.8	344	1 Y240_MYCPN	P75442 mycoplasma
5	92.5	4.5	513	1 TYRR_ECOLI	P07604 escherichia
6	92	4.5	812	1 MED_BUCAL	P57381 buchnera ap
7	91.5	4.5	388	1 HEMZ_STNEL	08d906 synchococc
8	87	4.3	1089	1 K406_HUMAN	043156 homo sapien
9	85.5	4.3	435	1 PIGU_HUMAN	09h490 homo sapien
10	85.5	4.2	463	1 VASI_MOUSE	09r199 mus musculu
11	85.5	4.2	974	1 SIN4_YEAST	P32259 saccharomyc
12	85	4.2	742	1 COC7_DROME	09v4d6 drosophila
13	85	4.2	844	1 APE2_YEAST	P32454 saccharomyc
14	84	4.1	1813	1 UN13_CABEL	P27715 caenorhabdi
15	83.5	4.1	434	1 P4G2_MOUSE	08k358 mus musculu
16	83.5	4.1	960	1 CAP2_MESCR	P16097 mesembryant
17	83	4.1	506	1 MATR_LATTI	08mc87 latyrus tl
18	83	4.1	513	1 TYRR_SALTY	054427 salmonella
19	83	4.1	655	1 SYT1_CANCR	09aa88 caulobacter
20	83	4.1	1960	1 MY9B_RAT	063358 rattus norv
21	82.5	4.0	284	1 ARLX_THENE	09z4s3 thermotoga
22	82.5	4.0	419	1 NOEI_RHISN	P55472 rhizobium s
23	82.5	4.0	463	1 VASI_RAT	054715 rattus norv
24	82	4.0	359	1 TP5H_CABEL	020351 caenorhabdi
25	82	4.0	513	1 MATK_MOLCA	09mu90 mollinia cae
26	82	4.0	1935	1 YBBA_SCHPO	06u072 schizosacch
27	82	4.0	2114	1 MY9B_MOUSE	09qy06 mus musculu
28	82	4.0	2337	1 TOR2_SCHPO	09y7k2 schizosacch
29	81.5	4.0	748	1 C184_NEUCR	042637 neurospora
30	81	4.0	320	1 F4SF_FLACH	P52837 flavaria ch
31	81	4.0	432	1 ORC4_XENLA	093479 xenopus lae
32	80.5	3.9	968	1 SNXD_HUMAN	09y5a8 homo sapien
33	80	3.9	429	1 THRC_SERMA	P27735 serratia ma

34	80	3.9	519	1 CPT7_RANDY	057525 rana dybows
35	80	3.9	1508	1 A10A_MOUSE	054827 mus musculu
36	79.5	3.9	531	1 UGTE_CABEL	010941 caenorhabdi
37	79.5	3.9	588	1 PESC_HUMAN	000541 homo sapien
38	79.5	3.9	966	1 CAP2_FLATR	P30694 flavaria tr
39	79.5	3.9	966	1 CAP2_FLATU	042730 flavaria au
40	79.5	3.9	1361	1 YME3_YEAST	004693 saccharomyc
41	79	3.9	495	1 ATPB_BEARE	09lmv0 beaucarnea
42	79	3.9	503	1 CP19_RAT	P51538 rattus norv
43	79	3.9	511	1 MATK_ORXSA	P12175 oryza sativ
44	79	3.9	513	1 MATK_PHRAU	09tci1 phragmites
45	79	3.9	1034	1 CND3_XENLA	09yhb5 xenopus lae

ALIGNMENTS

RESULT 1	ID	CGST_CHICK	STANDARD	PRT	458 AA.
AC	092179;				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	Chondroitin 6-sulfotransferase (EC 2.8.2.17) (C6ST).				
OS	Gallus gallus (Chicken).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;				
OC	Gallus.				
OX	NCBI_TaxID=9031;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 38-44; 46-52 AND 186-192.				
RC	STRATIN-White leghem; TISSUE-Embryonic chondrocytes;				
RX	MEDLINE=9535490; PubMed=7629189;				
RA	Fukuta M., Uchimura K., Nakashima K., Kato M., Kimata K.,				
RA	Shinomura T., Habuchi O.;				
RT	"Molecular cloning and expression of chick chondrocyte chondroitin 6-				
RT	sulfotransferase".				
RL	J. Biol. Chem. 270:18575-18580(1995).				
CC	- FUNCTION: CATALYZES THE SULFATION OF CHONDROITIN AND KERATAN				
CC	SULFATE.				
CC	- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + chondroitin -				
CC	adenosine 3',5'-bisphosphate + chondroitin 6'-sulfate.				
CC	- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI MEMBRANE				
CC	(BY SIMILARITY).				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; D49915; BAA08655.1; .				
DR	PIR; A57397; A57397.				
DR	InterPro: IPR000863; Sulfotransferase.				
DR	PIfam: PF00685; Sulfotransferase.				
KW	Transferrase; Transmembrane; Signal-anchor; Glycoprotein; Golgi stack.				
FT	DOMAIN	1	23		
FT	TRANSMEM	24	37		
FT	DOMAIN	38	458		
FT	CARBOHYD	62	62		
FT	CARBOHYD	73	73		
FT	CARBOHYD	95	95		
FT	CARBOHYD	236	236		
FT	CARBOHYD	399	399		
FT	CARBOHYD	443	443		
SO	SEQUENCE	458 AA;	52253 MM;	C9A3B7D0A508660C CRC64;	

Query Match 24.3%; Score 496.5; DB 1; Length 458;
Best Local Similarity 32.7%; Pred. No. 1,3e-35;

Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portelle D., Portolillo S., Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S., Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F., Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B., Sorokin A., Taccout E., Takagi T., Takahashi H., Takemaru K., Takeda M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A., Toso V., Uchiyama S., Vandembol M., Vannier F., Vassaret A., Viati A., Wampitt R., Wedler E., Wedler H., Wetzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zimstein E., Yoshikawa H., Zanchin A., "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*." Nature 390:249-256(1997).

RT Nature 390:249-256(1997).

RL [2]

RP SEQUENCE OF 1-1763 FROM N.A.

RC STRAIN-168 / PB1424:

RA Tognoni A., Grandi G.,

RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN SECONDARY METABOLISM.

CC -1- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES (POTENTIAL).

CC -1- SIMILARITY: Contains 4 acyl carrier domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: Z59113; CAB13603.1; -

DR EMBL: Z51133; CAB13603.1; -

DR PIR: C69679; C69679.

DR Subtilast; BG10931; pksM.

DR InterPro: IPR002198; ADH_short.

DR InterPro: IPR000794; Ketoacyl-synt.

DR InterPro: IPR001601; Methyltransf.

DR InterPro: IPR006163; pp_bind.

DR InterPro: IPR00051; SAM_bind.

DR Pfam: PF00106; adh_short; 1.

DR Pfam: PF00109; ketoacyl-synt; 3.

DR Pfam: PF00550; pp-binding; 4.

DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 2.

DR PROSITE: PS00606; B_KETOACYL-SYNTHASE; 2.

DR PROSITE: PS50075; ACP_DOMAIN; 4.

KW Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.

KM Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.

FT DOMAIN 295 364

FT DOMAIN 396 834

FT DOMAIN 2190 2258

FT DOMAIN 2322 2737

FT DOMAIN 3532 3947

FT DOMAIN 3410 3483

FT DOMAIN 4140 4209

FT BINDING 327 327

FT BINDING 2222 2222

FT BINDING 2476 2476

FT BINDING 3446 3446

FT ACT_SITE 3690 3690

FT BINDING 4172 4172

FT BINDING 103 103

FT CONFLICT 276 276

Q -> E (IN REF. 2).

FT CONFLICT 289 289 T -> S (IN REF. 2).

SO SEQUENCE 4273 AA; 477459 MM; 3BBFCFLA250AB5A CRC64;

Query Match 5.0%; Score 101.5; DB 1; Length 4273;

Best Local Similarity 22.3%; Pred. No. 5.3;

Matches 59; Conservative 44; Mismatches 99; Indels 63; Gaps 14;

QY 153 LUGGQPFPMVERACSHGFVVLKEVRFSLALYPLDLPDS---LNLHYVHLVNDPRA 208

DB 25 LPLGLAYIDLITQVDFEHY-AYQELKNTLITFYPLIDESTIALTHV----- 74

QY 209 VRSRHTTIELVDSHIVLQHLRTKE-----EDQPYAMRIICKSQYDIK 257

DB 75 --SEREGWSTIIDQ---KQGESLSKQRYETADMRKEQTAFASIDLNQKSTAD 129

QY 258 AITQLEPAQQRFLFRYEDLYRAPLAQTT-RYK-----FVGLDFPLHQTWYN 307

DB 130 RLINLDEIYEQ---CRSQELVHTGMKAKGQYEAKEGAVIDLAVGOALRHSDFLPH 185

QY 308 VT--RGKGQGHAFHTNARNALNVSAWMSLPEKVS--OLQDAGCEAMDLLGYLQVR 362

DB 186 PFLIDSGSIGS-----SCLISDQTMPLPLYESFSASERLQKC--TARILS-SSVR 234

QY 363 SQEQDGNLSIDLSSSHILGOVFE 387

DB 235 OKKELTYMTIEYFNSA---GQKVAE 256

RESULT 4

ID Y240_MYCPN STANDARD; PRT; 344 AA.

AC P75442;

DT 01-NOV-1997 (Rel. 35, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical protein MG240 homolog (F10_07f291).

GN MFN336 OR MF500.

OS Mycoplasma pneumoniae.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI_Taxid=2104;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 29342 / M129;

RX MEDLINE-97105885; PubMed-8948633;

RA Himmelfrich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,

RA Hermann R.;

RT "Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*."

RT Nucleic Acids Res. 24:4420-4449(1996).

RN [2]

RP REVISIONS.

RX MEDLINE-20411492; PubMed-10954595;

RA Dandekar T., Hyuen M., Regula J.T., Ueberle B., Zimmermann C.U.,

RA Andrade M.A., Doerks T., Sanchez-Pulido L., Snel B., Suyama M.,

RA Yuan Y.P., Hermann R., Bork P.;

RT "Re-annotating the *Mycoplasma pneumoniae* genome sequence: adding RT value, function and reading frames."

RT Nucleic Acids Res. 28:3278-3288(2000).

RL -1- SIMILARITY: TO B.SUBTILIS YOEK.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AE000049; AAB96148.2; -

CC HAMAP: MF_00244; atypical/fused; 1.

DR InterPro: IPR005249; Cons_hypoth488.

DR InterPro: IPR006674; HD.

DR InterPro: IPR003607; Met_phosphohydro.

CC -1- SIMILARITY: Contains 1 sigma-54 factor interaction ATP-binding
 CC domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M12114; AAA24706.1; -
 DR EMBL: AE000230; AAC74405.1; -
 DR EMBL: D90770; BAA14905.1; -
 DR EMBL: D90771; BAA14916.1; -
 DR PIR: A47086; RGECAV.
 DR Ecogene: Egl1042; tyrR.
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR002912; ACT.
 DR InterPro: IPR002197; HTH_Fls.
 DR InterPro: IPR000014; PAS_domain.
 DR InterPro: IPR002078; Sig54_Interact.
 DR Pfam: PF01842; ACT. 1.
 DR Pfam: PF00158; Sigma54_activat; 1.
 DR SMART: SM00382; AAA; 1.
 DR SMART: SM00091; PAS; 1.
 DR TIGRFAMs: TIGR01199; HTH_Fls; 1.
 DR PROSITE: PS50112; PAS; 1.
 DR PROSITE: PS00675; SIGMA54_INTERACT_1; 1.
 DR PROSITE: PS00676; SIGMA54_INTERACT_2; 1.
 DR PROSITE: PS00688; SIGMA54_INTERACT_3; 1.
 DR PROSITE: PS50045; SIGMA54_INTERACT_4; 1.
 KW Transcription regulation; Aromatic hydrocarbons catabolism;
 KW DNA-binding; Repressor; Activator; ATP-binding; Complete proteome.
 FT DOMAIN 78 149 PAS.
 FT NP_BIND 206 428 SIGMA-54 FACTOR INTERACTION (POTENTIAL).
 FT NP_BIND 234 241 ATP (POTENTIAL).
 FT NP_BIND 290 299 ATP (POTENTIAL).
 FT DNA_BIND 483 502 H-T-H MOTIF (POTENTIAL).
 SQ SEQUENCE 513 AA; 57656 MW; B580A401CA27866 CRC64;

Query Match 4.58; Score 92.5; DB 1; Length 513;

Best Local Similarity 19.48; Pred. No. 2.1; Matches 84; Conservative 71; Mismatches 157; Indels 121; Gaps 21;

CC 5 KKRILMFLGSOYIVVALFTIMSVYRHLSQRESRVRVHVLVLSMRSSSPFGOLGQH 64
 DB 116 KDRLRNHTAQAOLINGNFL-----RWLESEPODSHNEHVI-----NGQFLEI----- 161
 QY 65 PDVFYLMERPMHVMFTSTSTANKLHMAVR--DLRSVFLCDMSVFDAYMNPGRKOSL 122
 DB 162 TPVYLDQENDQHV---LTGAVVYLNRSTIMRQLGVAAQDVSAFQIYAVSP--KKKHV 216
 QY 123 FQWQDSALCSAYC-----DFF--PAHEISSPK-----HCKLLCGQPFEDVEKAC 167
 DB 217 VEQAQKLAISAPLLITGDTGKDLFAYACHQ--ASPRACKPYLALCASIPEDAVSEL 275
 QY 168 RSH-----GFVYLFKVRFLISQALYPLLTDPSSINHHVHLVRDPAVRSR 213
 DB 276 FGAPREGKGFEEQANGSVLLDDEIGMS-----PRMAKLLRFIND--GTER-- 321
 QY 214 EHTTIELMDSHVLGQHTETIKEDQPYAMKLIICKSQYDIYAKIOTLPEALQOQRLFL 273
 DB 322 -----RVEDGHEVHDVAVICATQKNIVELVQ--KGMRELLY 358
 QY 274 RYEDLV--RAPLAQTRLYFVGLDPLHQLTVYVNTKSGKGQGAFTNARNALNVSQ 331
 DB 359 RLNVLTLPRLDPCQ-----DIMPTELFFARFADQGVPRPLADLMTVLT--- 408
 QY 332 AMWMSLPYEKVSOLDAQGEAMDLLGLQYRSQ-----QOQNLSDLLS 376
 DB 409 --RYAMP--GNVRQKNAIYALITQLODYEELRPDILLPDYDAATVAVGEMSGSDEIT 465

QY 377 S---SHILGOVER 386
 DB 466 SRFERSVLTQLYR 478

RESULT 6
 MFD_BUCAI STANDARD; PRT; 812 AA.
 ID PF5381; 09F456;
 AC 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcription-repair coupling factor (TRCF).
 GN MFP OR BU294.
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
 OS symbiotic bacterium).
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Buchnera.
 OX NCBI_TaxID=118099;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Tokyo 1998;
 RX MEDLINE=20445173; PubMed=10993077;
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
 RT "Genome sequence of the endocellular bacterial symbiont of aphids
 RT Buchnera sp. Aps.,"
 RL Nature 407:81-86(2000).
 CC -1- FUNCTION: NECESSARY FOR STRAND-SPECIFIC REPAIR. A LESION IN THE
 CC TEMPLATE STRAND BLOCKS THE RNA POLYMERASE COMPLEX (RNAP) (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE UVRB FAMILY.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE HELICASE
 CC FAMILY. RECG SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AP001118; BAB13004.1; -
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR005118; TRCF.
 DR Pfam: PF02559; Card_TRCF; 1.
 DR Pfam: PF00270; DEAD; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF03461; TRCF; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC_C; 1.
 KW Helicase; DNA repair; ATP-binding; DNA-binding; Complete proteome.
 FT DOMAIN 263 633 HELICASE; RECG-LIKE.
 FT NP_BIND 293 300 ATP (POTENTIAL).
 FT SITE 394 397 DEAD BOX.
 SQ SEQUENCE 812 AA; 93881 MW; BB5C9F549EE2C35 CRC64;

Query Match 4.58; Score 92; DB 1; Length 812;

Best Local Similarity 17.88; Pred. No. 4.1;

Matches 63; Conservative 74; Mismatches 140; Indels 76; Gaps 16;

CC 41 PVHVLVSSMRSSSPFGOLGQHPDVFYLMERPMHVMFTSTANKLHMAVNDLRSV 100
 DB 206 PLHLGDDDMNKKKHKISKTYVDH-----AAQLHTYAKRESKTG-----A 247
 QY 101 FLCDMSVFDAYMNPGRKOSLFLQWQDSALCSAPVCDFFPAHEISSP--KXCKLLCGQPF 159
 DB 248 FKKNIEKYDLFCNDCKSRTS-----DQNEVW-----KVLKMSKRPIMPDRILCQDVG 296
 QY 160 FDNVEKACRSHGFVYLFKVRFLISQALYPLLTDP-----SLNHHV---HLVR 204
 DB 297 FGKTEIMARA--SELAVSNNKQVALIVPTTLAQOHKYNKFKIRFSNMPVNIINILSRFOTOK 355

RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Cervix, and Testis;
 RX MEDLINE-22388257; PubMed-12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Canninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.R., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fehey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC TISSUE SPECIFICITY: Widely expressed.
 CC -1- CAUTION: Ref.2 (CAC34556) sequence differs from that shown due to
 erroneous gene model prediction.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 DR EMBL: AB007866; BAA23702.2; ALT_INIT.
 DR EMBL: AL109823; CAC01531.2; ALT_INIT.
 DR EMBL: AL109823; CAC34556.1; ALT_SEQ.
 DR EMBL: BC013121; AH13121.1;
 DR EMBL: BC013755; AH13755.1;
 KW Polymorphism.
 FT DOMAIN 833 836 POLY-GLU.
 FT VARIANT 671 671 A -> V (IN DBSNP:1057238).
 FT VARIANT 979 979 A -> T (IN DBSNP:1064275).
 FT SEQUENCE 1089 AA; 122069 MW; 27F4D5EFA18A54E CRC64;
 SQ
 Query Match 4.3%; Score 88; DB 1; Length 1089;
 Best Local Similarity 19.9%; Pred. No. 13;
 Matches 83; Conservative 65; Mismatches 148; Indels 122; Gaps 21;
 QY 1 MMLKGRGLMFGSIVYVAFPHMSVHRHLSQRESRPVRVYLVLSWRSSGS--FVG 58
 DB 486 MLRQVCQLLGYGNGINYLVDHFMEL-YHOSVYRKAQ-----AMILNELVTGAAGLEVE 539
 QY 59 QLFQGH-----PDVFYLMERPAMHVMYMTSTANKL-----HNAVRLRLR 98
 DB 540 DLHEKHKTNP--ELKEIVTSTLEETYSQENMYLVYTCLETEMEGEELMHGLOQITS 597
 QY 99 SVFLCDMSVDFAYMNPGR--KRQSLFQ-----WEQSRALCSAPVQDFPAH 143
 DB 598 GEHTCQVTSFLASKSPTSCMSNSNTWCICLEGIGQFAVYAKGRFCLLSALYPLV 657
 QY 144 EISSPHKCLLGGQDPDWEKACRSHGVYLVKEVRFSLQALYPLITDPSLNLVHVHY 203
 DB 658 EKAGDQ--TLISQVATSTMDVCRACGYDSLQHL--INQNSDY-LVNGISLNLNR--HLA 710
 QY 204 RQD-----RAVRSREHTIIEIMVDSHYVLGQHLETKREDQPYVMKTKCSQNDVIY 257
 DB 711 LPHHTKRVLEVMRLNSDANILPLVAD--VVQDVLATL--DQFY-----DKRAMSEVS 758

QY 258 AIGTLPALQOQRYLFERYEDLVVRAPLAQTTRLYKFGVGLDPLPHIQVWYVNTNRGKMGCH 317
 DB 759 VLRHMLMALAQ-----WFPTGNNGHQ-----QSLGE 788
 QY 318 AFTNAPNALNVSOAMRWSLPEYKESQLODACEAMD---LGLYQVRSQOEGCNIS 371
 DB 789 GSHLNQRP-----ALEKSTTAEDIEQFLNLYLKEKDVAD-GNVS 828
 RESULT 9
 PIGU_HUMAN STANDARD; PRT; 435 AA.
 AC Q9H490; Q8N2F2;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cell division cycle protein 91-like 1 (CDC91-like 1 protein) (PIG-U).
 GN CDC91L1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Embryo;
 RA Hong Y., Kinoshita T.;
 RT "Human PIG-U and yeast Cdc91p are fifth subunits of GPI
 transamidase";
 RL submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-21638749; PubMed-11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.R., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.J., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Laylor S.,
 RA Leheeslath M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S., McConachie L.J., McElay K., McMurtry A.A.,
 RA Olive S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin R., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whitteker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilmshurst L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20";
 RL Nature 414:865-871(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Embryo;
 RA Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayaashi K., Ishii S.,
 RA Saito K., Yamamoto J., Wakamatsu A., Nagai T., Nakamura Y.,
 RA Nagahari K., Sugano S., Isogai T.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RX MEDLINE-22388257; PubMed-12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marasina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stajich M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carlino P., Prange C.,
 RA Rabe S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Boeck S.A., McKernan P.J., McKernan K.J., Malek J.A., Gamarale P.H.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shechenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scheraga A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: Belongs to the CDC91 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL, AB086842; BACS1626.1; -
 DR EMBL, AL185520; CAC14080.1; -
 DR EMBL, AK075507; BAC11660.1; -
 DR EMBL, BC030512; AAH30512.1; -
 DR Gene, HGNC:15791; CDC91L1.
 DR KW Transmembrane.
 FT TRANSMEM 1 21 POTENTIAL.
 FT TRANSMEM 66 86 POTENTIAL.
 FT TRANSMEM 166 186 POTENTIAL.
 FT TRANSMEM 188 208 POTENTIAL.
 FT TRANSMEM 237 257 POTENTIAL.
 FT TRANSMEM 259 279 POTENTIAL.
 FT TRANSMEM 286 306 POTENTIAL.
 FT TRANSMEM 313 333 POTENTIAL.
 FT TRANSMEM 355 375 POTENTIAL.
 FT TRANSMEM 386 406 POTENTIAL.
 FT CONFLICT 407 407 Y -> N (IN REF. 3).
 SQ SEQUENCE 435 AA; 50051 MW; 65FDD/96AB533027 CRC64;
 Query Match 4.3%; Score 87; DB 1; Length 435;
 Best Local Similarity 23.9%; Pred. No. 5;
 Matches 52; Conservative 31; Mismatches 73; Indels 62; Gaps 10;
 QY 9 LLMELGSOVIVVAFIHMVSRHLSQRESRPPVHVLSSMR-----SGSSFV 57
 DB 5 LVLVAVAVTVAALFRSLAEFISERVEVSP-----LSMKRYVEGLSLDLGVSPYS 59
 QY 58 GQLFQGHQDVY-----LMEPRMHWMTFTSSAMKLHMAVRDLASVFLCDMSVDAVYN 113
 DB 60 GAVFETPLITYLHFLIDYALVEMTDALTAIALYAIODFNKRVF----- 107
 QY 114 FGRFOSSILFQWEGSRALCSAP-VCDFF--PAHEISSPKHCKLGGQDFDWEKACRSH 170
 DB 108 ---KKQKLLLELDQ-----YAPDAVELIRTPMEKRYIPKLKALFLLNPTYLILSCVANKST 159
 QY 171 -----GFVVLKEVR-----FLSL--QALYPL 189
 DB 160 CAINTLLAFILITIKGSAPLSAFLALATATYQSLYPL 197
 RESULT 10
 VAS1_MOUSE STANDARD; PRT; 463 AA.
 AC Q9RIQ9;

DT 16-OCT-2001 (Rel. 40; Created)
 DT 16-OCT-2001 (Rel. 40; Last sequence update)
 DT 26-FEB-2003 (Rel. 41; Last annotation update)
 DE Vacuolar ATP synthase subunit S1 precursor (EC 3.6.3.14) (V-ATPase S1
 DE subunit) (V-ATPase S1 accessory protein) (V-ATPase Aca5 subunit) (C7-1
 DE protein).
 GN ATP6P1 OR ATP6S1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hayashi A., Hattori A., Okaze H., Kozuma S., Seki N., Saito T.,
 RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: VACUOLAR ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
 CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) -> ADP + phosphate +
 CC H(+) (out).
 CC -1- SUBUNIT: COMPOSED OF AT LEAST 10 SUBUNITS.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Vacuolar (By
 CC similarity).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL, AB031290; BAB83498.1; -
 DR MGD; MGI:109629; At6p1p1.
 DR KW ATP synthetase; Hydrogen ion transport; Hydrolase; ATP-binding;
 KW Transmembrane; Glycoprotein; Signal.
 FT SIGNAL 1 32 POTENTIAL.
 FT CHAIN 33 463 VACUOLAR ATP SYNTHASE SUBUNIT S1.
 FT TRANSMEM 413 433 POTENTIAL.
 FT CARBOHYD 164 164 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 255 255 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 267 267 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 290 290 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 297 297 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 344 344 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 351 351 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 359 359 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 463 AA; 51007 MW; AE28D99718BA0AC0 CRC64;
 Query Match 4.2%; Score 85.5; DB 1; Length 463;
 Best Local Similarity 23.4%; Pred. No. 7.4; Length 463;
 Matches 88; Conservative 33; Mismatches 112; Indels 143; Gaps 22;
 QY 89 LHMAYRDLRLSVFLCD-MSV--FDAYMMPGPKQSSLFQWEGSRALCSAPVCDFFPA-- 142
 DB 76 LELGRNVLV--LELDKLSIEDFTAYGVFGKQDSAPS--NLENMLDLAPSLVPAVDW 132
 QY 143 HEIS-----SPKCKLGGQDFDWEKACRSHGVYLVKEVFLSLQALYPL 190
 DB 133 YAISSLTYLLOEKLSAPLHVDL-----ATLKE--LKLNSLPAL 170
 QY 191 TDPSLNLHVHLYRDPRAVFRSREHTTIELWDSHIVLGOMLETKEEDQY-VAMKTIIC 249
 DB 171 L-----LIRLPYTA--SSGLMAPREVLTGNDVETIGVYSLTKSEVPTALATVVR 219
 QY 250 KSOV--DIVKAIOTLPALQO-----RYLF-----LRYED----- 277
 DB 220 PSRVARDITWAGGLGRLQOVASPAIHPPVSYNDATPRILFPAQNFSAVAKQEMKDL 279
 QY 278 -----LVRAPIAQTTIRLYKFV-GIDFLP-HIQTVV----- 305
 DB 280 TSLFGEVNLNTGSSFMWDSFAMLSLTYEPLGATVTEFKILASRFYPSARVYPMABRL 339
 QY 306 -----YVYTKGKMGQHAHF-----TNARNAL--NVSQAMKMSLPYEVKSQLO 346

DB 340 EHSNCSVAHFNVSOVTPSYFHCYVSSVSKGNLVTNVPSVMQMLHNPQI-QAF 398
 QY 347 DAGE-----AMDLG 358
 DB 399 NVTGEQFSVSDCAF 414

RESULT 11
 SIM_YEAST
 ID SIM_YEAST STANDARD; PRT; 974 AA.
 AC P32259;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Global transcriptional regulator SIM4.
 GN SIM4 OR TSF3 OR BEL2 OR GAL22 OR SSF5 OR YNL26W OR N1135.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YM256;
 RX MEDLINE=93140781; PubMed=8423805;
 RA Chen S., West R.W. Jr., Johnson S.L., Gans H., Kruger B., Ma J.;
 RT "TSF3, a global regulatory protein that silences transcription of
 RT yeast GAL genes, also mediates repression by alpha 2 repressor and is
 RT identical to SIM4.";
 RL Mol. Cell. Biol. 13:831-840(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93024394; PubMed=1406639;
 RA Jiang Y.W., Stillman D.J.;
 RT "Involvement of the SIM4 global transcriptional regulator in the
 RT chromatin structure of Saccharomyces cerevisiae.";
 RL Mol. Cell. Biol. 12:4503-4514(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Harashima S., Mabuuchi H., Ramash R., Hasebe M., Tanaka A., Oshima Y.;
 RL Submitted (MUG-1992) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97051596; PubMed=8896273;
 RA Pandolfo D., de Antoni A., Lanfranchi G., Valle G.;
 RT "The DNA sequence of cosmid 14-5 from chromosome XIV reveals 21 open
 RT reading frames including a novel gene encoding a globin-like
 RT domain.";
 RL Yeast 12:1071-1076(1996).
 CC -1- FUNCTION: GLOBAL REGULATORY PROTEIN THAT SILENCES TRANSCRIPTION
 CC OF GAL AND MATING-TYPE GENES. NEGATIVE REGULATOR OF THE HO
 CC (HOMOTHALISM) GENE. MAY POTENTIATE TRANSCRIPTIONAL ACTIVATION
 CC AND REPRESSION BY REGULATING THE ACTIVITY OF THE BASAL
 CC TRANSCRIPTIONAL APPARATUS. SIM4 MUTATION ALTERS CHROMATIN
 CC STRUCTURE.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M93050; AAA35044.1; -
 CC EMBL: X64516; CAA45819.1; -
 CC EMBL: D12918; BAA02302.1; -
 CC EMBL: Z69381; CAA93362.1; -
 CC EMBL: Z71512; CAA96140.1; -
 CC PIR: A44484; A44484.
 CC TRANSFAC: T01243;
 CC SGD: S0005180; SIM4.

DR GO: GO:0000119; C:mediator complex; IDA.
 DR GO: GO:0016455; F:RNA polymerase II transcription mediator ac. . . IDA.
 DR GO: GO:0006366; P:transcription from pol II promoter; IDA.
 KW Transcription regulation; Activator; DNA-binding; Repressor;
 KW Nuclear protein.
 FT DOMAIN 63 81 SER/THR-RICH.
 FT DOMAIN 889 893 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT SEQUENCE 974 AA; 111296 MW; 12ECF5EACDC05A8E CRC64;
 SO Query Match 4.28; Score 85.5; DB 1; Length 974;
 Best Local Similarity 17.7%; Pred. No. 19;
 Matches 65; Conservative 49; Mismatches 131; Indels 123; Gaps 15;

QY 24 IHMSVHRLSORESRPRPVHVLSSRRSSFPGL- -GGQ- - - - -POVF-LMPPAW 75
 DB 51 LHPOLHEVOYOEBSSTLSTSTTSTVNGSTAGVSTPNNGSNKSPQPFYNISSHW 110
 QY 76 HWMTF-----TSSWAKLHMAVRDLRSVFLCDMSVDAY 111
 DB 111 NNMFSLPGDMLAVCDELGNMTMLTIGRPRATTYEKLTVFPD- - - - -NVYKIY 160
 QY 112 MNPP- - - - -KRSSL- - - - -FGW-EQSRALCAPYCD- - - - - 139
 DB 161 NHVPLRPVDPKLPNMNERQTRKEYNTSLIEFWLTSKSVISQFCARDSSNTYRSR 220
 QY 140 - - -FPAREISPKHCKLLC- - - - -GOQPPDMWEKACRSHGFVL- - - - -KEVRF 180
 DB 221 AQOVPPYGVYHPPIKACLAIRKNGQIDFWQFSNKRDKKTKTLQLLDTSNOREKDLQW 280
 QY 181 LSLQALYPLLTDPSSL- - - - -NLHV- - - - -VHLVDPRAVRSREHTT 217
 DB 281 LEFARITPMNDQCMILITTSYSKLSKNISFYKLHVMNLNMPVNLNDPSLKIQFILSTT 340
 QY 218 IELMVD-SHYLGCHLETI- - - - -KEEDPYAMKTIKCSQVDIVAIQTLPBALQRY 270
 DB 341 LDPTDGHVJLKENLHVSKSSTKDPSEIILVYVCDTSKSLVRYRLAPQLSAEY 400
 QY 271 LFLRYEDL 278
 DB 401 LVILKPOL 408

RESULT 12
 COG7_DROME
 ID COG7_DROME STANDARD; PRT; 742 AA.
 AC OGYAD6; O8SZV9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Putative conserved oligomeric Golgi complex component 7.
 GN CG31040/CG7880.
 OS Drosophila melanogaster (Fruit fly)
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams R.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borkov D., Botchan M.R., Bouck B.P., Brockstein P., Brotter P.,
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibeigbam C.,
 RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kension J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merulov G., Misha N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman K.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sildenafil A.C., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of Drosophila melanogaster.";
 RA Science 287:2185-2195(2000).
 RL [2]

RP REVISIONS.
 RC STRAIN-Berkeley; PubMed-12537572;
 RA MEDLINE-22426069; PubMed-12537569;
 RA Mista S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.B., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Bernan B.P.,
 RA Bettecourt B.R., Celinker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RA "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review.";
 RT Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 RL [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN-Berkeley; TISSUE-Embryo;
 RA MEDLINE-22426066; PubMed-12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celinker S.E.;
 RA "A Drosophila full-length cDNA resource.";
 RT Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
 CC -1- FUNCTION: Required for normal Golgi function (by similarity).
 CC Is composed of eight different subunits and is required for normal
 CC Golgi morphology and localization (by similarity).
 CC -1- SUBUNIT: Component of the conserved oligomeric Golgi complex which
 CC is composed of eight different subunits and is required for normal
 CC Golgi morphology and localization (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE COG7 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, AE003771; AAF56973.2; -;
 DR EMBL, AY069758; AAL39903.1; -;
 DR HSSP, P32906; IDL2.
 DR FLBASE: P890051040; CG31040.
 DR Interpro: IPR001382; Glyco_hydro_47.
 DR Pfam: PF01532; Glyco_hydro_47; 1.
 DR PRINTS: PR00747; GLYHDLASE7.
 KW Hypothetical protein; Transport; Protein transport; Golgi stack;
 KW Membrane.
 SO SEQUENCE 742 AA; 84075 MW; 85998ER36519AE39 CRC64;

Query Match 4.2%; Score 85; DB 1; Length 742;
 Best Local Similarity 22.1%; Pred. No. 15;
 Matches 53; Conservative 40; Mismatches 99; Indels 48; Gaps 8;
 Oy 180 FLSIALQALYLLDPSLNHVALVR-----DPAVFRRETRTTT-----ELAVDSHIYVG 229
 Db 288 FLVIAELLPAL-QPFRDAHIIQLLTNSNERLEMLLFAKVNISFVLHNLNSLEOSHITLS 346
 Oy 230 QHLETFIKEDQPYAMKLT-----CRSQVDIVKAIQTLP-----EALQORYL 271
 Db 347 BELHNLGALAEYEFHFKFLQYPRLEFQLSTQVDRLSNQATPPDGVNHLSESTRKLYE 406
 Oy 272 FLRYEDLVAPLAQTRLYKFEVGL---DFPLHQLQTVYVNTVRGKMGQAHFTNARNALN 328
 Db 407 WLKFACEKCASTSDIALCKLTITLNLGFKRQLESF-----GRIGRQIGLSIGSSSY 458
 Oy 329 VSOAMRWSPYKRYQQLDQACGEMDLGYLQVRSQEQEGNLSIDLLSSHTLQGVFRG 388
 Db 459 AAQSENWSDLYQTYMSQLD-----CLADFQVQLHQPQDLHTRVYMLNRLTPSNRG 510
 RESULT 13
 APE2_YEAST STANDARD; PRT; 844 AA.
 ID APE2_YEAST
 AC P32454;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Amino-peptidase II (EC 3.4.11.-) (YscII).
 GN APE2 OR LAP1 OR YKL157W OR YKL611.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AB320;
 RA MEDLINE-92111552; PubMed-1765107;
 RA Garcia-Alvarez N., Cueva R., Suarez-Rendueles P.;
 RT "Molecular cloning of soluble amino-peptidases from Saccharomyces
 RT cerevisiae. Sequence analysis of amino-peptidase yscII, a putative
 RT zinc-metallo-peptidase.";
 RL Eur. J. Biochem. 202:993-1002(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RA MEDLINE-94378720; PubMed-8091859;
 RA Vandenbol M., Bolle P.-A., Dion C., Portetelle D., Hilger F.;
 RT "DNA sequencing of a 36.2 kb fragment located between the FAS1 and
 RT LAP loci of chromosome XI of Saccharomyces cerevisiae.";
 RL Yeast 10:535-540(1994).
 CC -1- FUNCTION: IT IS INVOLVED IN THE CELLULAR SUPPLY OF LEUCINE FROM
 CC EXTERNALLY OPERATED LEUCINE-CONTAINING DIPEPTIDE SUBSTRATES.
 CC -----
 CC -1- COFACTOR: Binds 1 zinc ion (by similarity).
 CC -1- SIMILARITY: Belongs to peptidase family M1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, X63998; CAA45403.1; -;
 DR EMBL, Z26877; CAA81497.1; -;
 DR EMBL, Z28157; CAA81999.1; -;
 DR PIR, S37794; S37794.
 DR MEROPS, M01.006; -;
 DR SCD, S0001640; APE2.
 DR GO: GO:0005737; C:cytoplasm; NAS.
 DR GO: GO:0030287; C:periplasmic space (sensu Fungi); NAS.

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 18:10:01 ; Search time 543 seconds
(without alignments)
10136.449 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043
Sequence: 1 gaattccattggtgtggtgta.....tgggatacttaaaaaaaaaa 2043

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2552756 segs, 1349719017 residues

Word size: 0

Total number of hits satisfying chosen parameters: 4193764

Minimum DB seq length: 15

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: N_Geneseq_19Jun03:*

```
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	1547	75.7	1979	22	AAK94229 Human full-length
2	1547	75.7	2032	20	AAZ20792 Human glycosyl sul
3	1262	61.8	1333	24	AA516947 Human l-selectin s
4	1153	56.4	2065	21	AAZ94211 Human transferrase
5	695	34.0	877	22	AAK91803 Human CDNA 5'-end
6	695	34.0	877	22	AAK93921 Human CDNA clone r
7	403	19.7	505	24	ABV89280 Human colon cancer
8	403	19.7	517	24	ABK54724 Human colon cancer

9	306	15.0	2988	21	AAK76156 Human OREF ORF1711
10	114	5.6	389	24	ABK54794 Human colon cancer
11	105	5.1	548	22	AAK92588 Human CDNA 3'-end
12	59	2.9	1647	24	AAK24670 Human drug metabol
13	59	2.9	1694	22	AAK2700 Human glycosyl sul
14	59	2.9	2044	22	AAK2699 Human glycosyl sul
15	59	2.9	2170	22	AAK2698 Human glycosyl sul
16	59	2.9	2544	24	ABN69505 Human corneal N-ac
17	59	2.9	48436	24	ABN69533 Human corneal N-ac
18	59	2.9	160552	22	AAK2697 Human glycosyl sul
19	37	1.8	37	20	AAK20798 Human glycosyl sul
20	30	1.5	30	24	ABK70821 PCR primer for gly
21	27	1.3	1926	20	AAK20793 Probe for human ge
22	23	1.1	1937	24	AAK16948 Murine intestinal-
23	23	1.1	1989	22	AAK20696 Mouse glycosyl sul
24	22	1.1	22	24	ABK70834 PCR primer for hum
25	22	1.1	22	24	AAK16960 Human l-selectin s
26	22	1.1	22	24	AAK16962 Human l-selectin s
27	22	1.1	22	24	AAK16964 Human l-selectin s
28	22	1.1	167	24	ABK6480 Human ovarian canc
29	22	1.1	171	24	ABK6089 Human ovarian canc
30	22	1.1	187	24	ABK6207 Human ovarian canc
31	22	1.1	191	24	ABK6306 Human ovarian canc
32	21	1.0	21	24	ABK70835 PCR primer for hum
33	21	1.0	21	24	AAK16961 Human l-selectin s
34	21	1.0	10004	22	ABK14483 Human nervous syst
35	21	1.0	10246	22	ABK14485 Human nervous syst
36	20	1.0	88	21	AAK25980 Human secreted pro
37	20	1.0	354	22	AAK65813 Novel human polynu
38	20	1.0	370	22	AAK65341 Novel human polynu
39	20	1.0	381	22	AAK67389 Bovine EST associa
40	20	1.0	398	25	ABK45934 Human myosin l-cha
41	20	1.0	3763	19	AAV58194 DNA encoding novel
42	20	1.0	3763	23	AAK569872 Human ovary specif
43	20	1.0	5814	24	ABN87855 Human dlthp protei
44	20	1.0	6147	25	ACC46453 Human immune/haema
45	20	1.0	9516	22	AAK73555

ALIGNMENTS

RESULT 1	AAK94229	standard; CDNA; 1979 BP.
ID	AAK94229	
AC	AAK94229;	
DT	06-NOV-2001	(first entry)
DE	Human full-length CDNA, SEQ ID NO: 2816.	
XX	Human; full length CDNA; CDNA synthesis; oligo-capping; ss.	
OS	Homo sapiens.	
PN	EP130094-A2.	
XX	05-SEP-2001.	
PD	07-JUL-2000; 2000EP-0114089.	
XX	08-JUL-1999; 99JP-0194486.	
XX	PR 11-JAN-2000; 2000JP-0118774.	
PR	02-MAY-2000; 2000JP-0183765.	
XX	(HELI-) HELIX RES INST.	
PA	Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;	
PI	Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;	
PI	WPI: 2001-524255/58.	
XX	P-PSDB; AAM93309.	
DR		

KW secondary lymph organ; ss.
 XX Homo sapiens.
 OS MO9949018-A1.
 XX 30-SEP-1999.
 PD
 XX 26-FEB-1999; 99WO-US04316.
 XX
 PR 20-MAR-1998; 98US-0045284.
 PR 12-NOV-1998; 98US-0190911.
 XX
 PA (REGC) UNIV CALIFORNIA.
 PA (SYNT) SYNTX USA INC.
 XX
 PI Blstrup A, Rosen SD, Tangemann K, Hemmerich S;
 DR WPL; 1999-580442/49.
 DR P-PSDB; AAY39918.
 XX
 PT Human and murine glycosyl sulfotransferase 3 and related
 XX polynucleotides
 PS
 XX Claim 4; Fig 1; 59pp; English.
 CC This sequence encodes the human glycosyl sulfotransferase-3 (GST-3) of
 CC the invention. The nucleic acid sequences, probes and primers derived
 CC from these, proteins and antibodies are useful in detecting homologues.
 CC The sequences, antibodies and methods are useful in the diagnosis and
 CC treatment of diseases associated with selectin binding interactions,
 CC including conditions associated with or resulting from the homing of
 CC leukocytes to sites of inflammation and the normal homing of lymphocytes
 CC to secondary lymph organs.
 CC
 SO Sequence 2032 BP; 468 A; 569 C; 490 G; 505 T; 0 other;
 Query Match 75.7%; Score 1547; DB 20; Length 2032;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 586 GCTCTGTCAGTCAACAGCCCTTTGAGTGTGTGAGAGAGCGTCCGCTCTACAGCA 645
 DB 640 GCTCTGTGAGTCAACAGCCCTTTGAGTGTGTGAGAGAGCGTCCGCTCTACAGCA 699
 QY 646 CGTGTGTCTAAGAGAGGTGCGCTTCTTCAACCTCGAGTCCCTTACCCTGCTGAAAGA 705
 DB 700 CGTGTGTCTAAGAGAGGTGCGCTTCTTCAACCTCGAGTCCCTTACCCTGCTGAAAGA 759
 QY 706 CCCCTCCCTCAACCTGCAATATGCTGACCTGTGCGGGAGCCCGGGCGCTGCTCC 765
 DB 760 CCCCTCCCTCAACCTGCAATATGCTGACCTGTGCGGGAGCCCGGGCGCTGCTCC 819
 QY 766 CCGAGAACGCAAGAGAGAGATCTCATATGATGACAGTGTGATGATGGGAGAGATGA 825
 DB 820 CCGAGAACGCAAGAGAGAGATCTCATATGATGACAGTGTGATGATGGGAGAGATGA 879
 QY 826 GCAAAAACCTCAAGAGAGAGACCAACCTTACTATGTGTGAGGTCAATCTGCCAAGCCA 885
 DB 880 GCAGAAACTCAAGAGAGAGAGACCAACCTTACTATGTGTGAGGTCAATCTGCCAAGCCA 939
 QY 886 GCTGGAGATCTACAAAGACATCCAGTCTCTGCCCAAGGCCCTCGAGAGACCTACTGCT 945
 DB 940 GCTGGAGATCTACAAAGACATCCAGTCTCTGCCCAAGGCCCTCGAGAGACCTACTGCT 999
 QY 946 TGTGCGCTATGAGAGACCTGCGTCAAGCCCTGTGGCCAGACTTCCCGAATGTAGATT 1005
 DB 1000 TGTGCGCTATGAGAGACCTGCGTCAAGCCCTGTGGCCAGACTTCCCGAATGTAGATT 1059
 QY 1006 CGTGGAGATGGAATTTCTTGCCCATCTTCAGACCTGGGTGATTAATCAATCACCAGGCA 1065
 DB 1060 CGTGGAGATGGAATTTCTTGCCCATCTTCAGACCTGGGTGATTAATCAATCACCAGGCA 1119
 QY 1066 GGGCATGGGTGACCAAGCTTTCACACAAAGGCAAGGATCCCTTATGTCTCCAGGC 1125
 DB 1120 GGGCATGGGTGACCAAGCTTTCACACAAAGGCAAGGATCCCTTATGTCTCCAGGC 1179
 QY 1126 TTGGCGCTGCTTTTGGCCCTATGAAAAGTTTCTGATTCAGAAAGCCTGTGGAGATC 1185
 DB 1180 TTGGCGCTGCTTTTGGCCCTATGAAAAGTTTCTGATTCAGAAAGCCTGTGGAGATC 1239
 QY 1186 CATGAATTTGCTGGGCTTACCGCCACAGTCTGATCTGAAACAGAGAAACCTGTGCT 1245
 DB 1240 CATGAATTTGCTGGGCTTACCGCCACAGTCTGATCTGAAACAGAGAAACCTGTGCT 1299
 QY 1246 GGATCTTGTCTACCTGAGATGTCCTGAGCAAAATCAGTAAGGGTTGAGAAAGCTT 1305
 DB 1300 GGATCTTGTCTACCTGAGATGTCCTGAGCAAAATCAGTAAGGGTTGAGAAAGCTT 1359
 QY 1306 TGTGCAACCTGTGTGACCTCAAGTCACTTCTCTGAATGCTTCTGAAGCTTGCTTACA 1365
 DB 1360 TGTGCAACCTGTGTGACCTCAAGTCACTTCTCTGAATGCTTCTGAAGCTTGCTTACA 1419
 QY 1366 TCTCTGAGCCTTAACTACATCTGTGTGGATATCACACTGAGTGTGATGCTCCACAG 1425
 DB 1420 TCTCTGAGCCTTAACTACATCTGTGTGGATATCACACTGAGTGTGATGCTCCACAG 1479
 QY 1426 TGTCTCAAGCAGAAGACTTGTGTGCTTGTGCTTGTGTGCTTGTGTGCTTGTGTGCT 1485
 DB 1480 TGTCTCAAGCAGAAGACTTGTGTGCTTGTGCTTGTGTGCTTGTGTGCTTGTGTGCT 1539
 QY 1486 TATGTGAGCAGACATCCACAGTGAAGAGGATATGCTTCTTCTTCTTCTTCTGATCT 1545
 DB 1540 TATGTGAGCAGACATCCACAGTGAAGAGGATATGCTTCTTCTTCTTCTTCTGATCT 1599
 QY 1546 TCTGTCTGGGAGACTTCAAGACTTGTGGCTGTGAGGCTTATTAAGCAGACAGT 1605
 DB 1600 TCTGTCTGGGAGACTTCAAGACTTGTGGCTGTGAGGCTTATTAAGCAGACAGT 1659
 QY 1606 ATCAGTGAATTTGATCCATTAATCCCTGTCACATCTTCCCAATGGGAAATGATCT 1665
 DB 1660 ATCAGTGAATTTGATCCATTAATCCCTGTCACATCTTCCCAATGGGAAATGATCT 1719
 QY 1666 TTCACAAAGAGCTCACAGCATTTTCCACAGAGATGC 1703

Db 1720 TTCACCAAGAGCTCACCAGCATTTCCACAGAGATGC 1757

RESULT 3

ID	AAS16947 standard; cDNA; 1333 BP.

AC AAS16947;

DT	12-MAR-2002	(first entry)
VV		

Human L-selectin sulfotransferase-2 (LSST-2) cDNA

KM Human; beta1,3Gnt; beta1,3-N-acetylglucosaminyltransferase; MEGA-79; ss
KM L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
KM ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;
KM allergic contact dermatitis; Lymphoma; chronic pneumonia; LST-2;
KM delayed-type hypersensitivity reaction; hyperplastic thymus; antitumor;
KM antiinflammatory; antipsoriatic; antidiabetic; dermatological;
KM antiallergic.

OS Homo sapiens.

FH	Key	Location/Qualifiers
ET	CDS	111 1253

FT	CD3	111...1233	/#tag= 3
FT			

FT /product= "Human LSST-2"

PN WO200185177-A1.

PD 15-NOV-2001

PF 10-MAY-2001; 2001WO-US15452.
VY

PR 11-MAY-2000; 2000US-0569320.
VY

PA (BURN-) BURNHAM INST.
XY

Pl Fukuda M, Yeh J, Hiraoka N,
XX

DR WPI; 2002-075226/10.
DR P-PSDB: AAPI1274

New enzyme: useful

isolated *V*-selectin sulfotransferase-2 that directs expression of *V*-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or intestinal GlcNAc 6-sulfotransferase

PS Claim 19; Flg 4; 98pp; English.

The present invention provides a method of modifying an acceptor molecule by contacting the acceptor with an isolated betaal,3-N-acetylglucosaminyltransferase (betaal,3GNTr) or an active fragment, where betaal,3GNTr directs expression of a MECA-79 antigen. The invention also provides a method of treating or preventing an I-selectin-mediated condition by reducing the expression or activity of a betaal,3GNTr that directs expression of a MECA-79 antigen. This can be done by administering to the subject an oligosaccharide I-selectin antagonist that inhibits binding of I-selectin to a MECA-79 antigen, for example by administering antibody material that specifically binds betaal,3GNTr, and/or a betaal,3GNTr antisense nucleic acid molecule. I-selectin sialyltransferase-2 (LSSr-2) also directs MECA-79 antigen expression. Alternatively, the expression or activity of LSSr-2 or its active fragment can be reduced in combination with reducing the expression or activity of betaal,3GNTr. The method is useful for treating I-selectin mediated conditions such as Crohn's disease and ulcerative colitis, inflammatory disorders of the skin such as allergic contact dermatitis, psoriasis and Lichen planus, lymphomas, chronic pneumonia, delayed-type hypersensitivity reactions, diabetes and hyperplastic thymus. This sequence represents cDNA encoding human LSSr-2.

50 Sequence 1333 BP; 286 A; 393 C; 336 G; 318 T; 0 other;

Query Match	61.8%;	Score 1262;	DB 24;	Length 1333;
Best Local Similarity	100.0%;	Pred. No. 0;	.	
Matches 1262;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;

OY	107	AAGGCTTCCACTTCAGACAAATGGTACTGGCTAAAAAAAAGAACTCTGCTGTTTTCG	166
Db	72	AAGGCTTCCACTTCAGACAAATGGTACTGGCTAAAAAAAAGAACTCTGCTGTTTTCG	131
OY	167	GTTCGCCAGATGCCATCTTGGCTCTATTTCTTCACATGTACAGCCACACATCAGCTCC	226
Db	132	GTTCGCCAGATGCCATCTTGGCTCTATTTCTTCACATGTACAGCCACACATCAGCTCC	191
OY	227	CTGTCTATGAAGGACACAGCCCGAGGGCATGTGACGGTGGTGTGTCTCTCCGGAGCTC	286
Db	192	CTGTCTATGAAGGACACAGCCCGAGGGCATGTGACGGTGGTGTGTCTCTCCGGAGCTC	251
OY	287	GGCTCTCTTTTGTGGGGCAGCTTTTGTGGGAGCACCCAGATGTTTTCTACCTGATGAG	346
Db	252	GGCTCTCTTTTGTGGGGCAGCTTTTGTGGGAGCACCCAGATGTTTTCTACCTGATGAG	311
OY	347	CCCGGCTTGGACGCTGTGATGACCTCTCAAGAGACACCGCTGGATGCTGCACATGGCT	406
Db	312	CCCGGCTTGGACGCTGTGATGACCTCTCAAGAGACACCGCTGGATGCTGCACATGGCT	371
OY	407	GTGCGGATCTGTATPACGGGGCGTCTTCTTGTGTGCACATGAGCGTCTTGTATGCTTAC	466
Db	372	GTGCGGATCTGTATPACGGGGCGTCTTCTTGTGTGCACATGAGCGTCTTGTATGCTTAC	431
OY	467	GAACCTGGTTCGCCGGAGACGTCCACAGCCCTCTTCAGTGGGGAACACGCGGGCCCTGTAT	526
Db	432	GAACCTGGTTCGCCGGAGACGTCCACAGCCCTCTTCAGTGGGGAACACGCGGGCCCTGTAT	491
OY	527	TCTGCACCTGCTGTGACATATCCACAGAATGAATCATCCCCGGGGCTCATCTGACGG	586
Db	492	TCTGCACCTGCTGTGACATATCCACAGAATGAATCATCCCCGGGGCTCATCTGACGG	551
OY	587	CTCCGCTGCATPACAAACAGCCCTTTGAGGTGGTGGGAAGGGCTGGCCGCTCTACAGCCAC	646
Db	552	CTCCGCTGCATPACAAACAGCCCTTTGAGGTGGTGGGAAGGGCTGGCCGCTCTACAGCCAC	611
OY	647	GTGCTGCTCAAGAGAGGTGCGGCTTCTTCAACCTGCAGTCCCTCTACCCGCTCTGAAAGAC	706
Db	612	GTGCTGCTCAAGAGAGGTGCGGCTTCTTCAACCTGCAGTCCCTCTACCCGCTCTGAAAGAC	671
OY	707	CCCTCCCTCAACCTGCATATGTGTGACCTGGTCCGGGGACCCCGGGGCCGTGTCCGTTCC	766
Db	672	CCCTCCCTCAACCTGCATATGTGTGACCTGGTCCGGGGACCCCGGGGCCGTGTGTCCGTTCC	731
OY	767	CGAGAGCCACAAAGGAGATCTCATGTTTACAGCTGCATTTGTGATGGGGCAGCATGAG	826
Db	732	CGAGAGCCACAAAGGAGATCTCATGTTTACAGCTGCATTTGTGATGGGGCAGCATGAG	791
OY	827	CAAAAACCTCAAGAAAGAGAGACCAACCCCTCATGTGATGAGGAGGTGATGTGCAAAAGCAG	886
Db	792	CAAAAACCTCAAGAAAGAGAGACCAACCCCTCATGTGATGAGGAGGTGATGTGCAAAAGCAG	851
OY	887	CTGAGATCTTCAAGAACCATCATGCTCTTGGCCCAAGGCCCTGCGAAGAAAGCTACTCTGTT	946
Db	852	CTGAGATCTTCAAGAACCATCATGCTCTTGGCCCAAGGCCCTGCGAAGAAAGCTACTCTGTT	911
OY	947	GTGGGCTATGAGAGACCTGGCTCGAGCCCTGTGGGCCACGATTCGCCGAATGTATGAATTC	1006
Db	912	GTGGGCTATGAGAGACCTGGCTCGAGCCCTGTGGGCCACGATTCGCCGAATGTATGAATTC	971
OY	1007	GTGGGATTTGAAATCTTGCCCCATCTTCAGACTGTGGGTGCAATTAATCAATCAACCCGAGC	1066
Db	972	GTGGGATTTGAAATCTTGCCCCATCTTCAGACTGTGGGTGCAATTAATCAATCAACCCGAGC	1031
OY	1067	GGCATGGGTGACACAGCTTTTCCACACAATAATGCCAGAGATGCCCTTAAATGTCTCCAGGCT	1126
Db	1032	GGCATGGGTGACACAGCTTTTCCACACAATAATGCCAGAGATGCCCTTAAATGTCTCCAGGCT	1091

QY 1127 TGGCGCTGCTTTTGGCCCTATGAAAGATTCTCGACTTCAGAAAGCCTGTGGGATGCC 1186
 Db 1092 TGGCGCTGCTTTTGGCCCTATGAAAGATTCTCGACTTCAGAAAGCCTGTGGGATGCC 1151
 QY 1187 ATGAATTTGCTGGCTTACCCGACAGTGTGATCTGACAGAACAGAAACCTGTGCTG 1246
 Db 1152 ATGAATTTGCTGGCTTACCCGACAGTGTGATCTGACAGAACAGAAACCTGTGCTG 1211
 QY 1247 GATCTTCTGCTACCTGAGCTGTGCTGACCAATTCAGTAAAGGCTTGAAGGCTTT 1306
 Db 1212 GATCTTCTGCTACCTGAGCTGTGCTGACCAATTCAGTAAAGGCTTGAAGGCTTT 1271
 QY 1307 GCTGCCACCTGCTGTGCTGAGCTGTGCTGATCTGCTGATGCTTGGCTACAT 1366
 Db 1272 GCTGCCACCTGCTGTGCTGAGCTGTGCTGATCTGCTGATGCTTGGCTACAT 1331
 QY 1367 CT 1368
 Db 1332 CT 1333

RESULT 4

AAZ94211
 ID AAZ94211 standard; cDNA; 2065 BP.

AC AAZ94211;

DT 19-JUN-2000 (first entry)

DE Human transferase TRNSFS-11 cDNA clone 2617407CBI.

XX Transferase; TRNSFS-11; human; antitumor; cell proliferation;

KM inflammation; gastrointestinal disorder; developmental disorder;

KM genetic disorder; neurological disorder; reproductive disorder;

KM smooth muscle disorder; immunological disorder; gene therapy;

KM diagnosis; N-acetylglicosamine 6-O-sulfotransferase; ss.

XX Homo sapiens.

OS Homo sapiens.

FT Key location/Qualifiers

FT CDS 174..1334

FT /*tag= a

PN MO200014251-A2.

XX 16-MAR-2000.

PD 09-SEP-1999; 99WO-US20989.

XX 10-SEP-1998; 98US-0150657.

PR 04-NOV-1998; 98US-0186779.

PR 11-MAY-1999; 99US-0133642.

XX (INCYTE) INCYTE PHARM INC.

XX Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;

PI Hillman JL, Azimzal Y;

XX WPI: 2000-256996/22.

DR P-PSDB; AAY79219.

XX Human transferase proteins useful for preventing, diagnosing and

PT treating cancers and developmental, gastrointestinal, genetic, and

PT immunological, neurological, reproductive and smooth muscle disorders -

XX Claim 9; Page 104-105; 113pp; English.

XX The present sequence is that of cDNA clone 2617407CBI encoding

CC human transferase TRNSFS-11 (see AAY79219). 1 of 15 claimed human

CC transferase proteins of the invention (see AAY79209-23). The clone

CC was isolated from gall bladder cDNA library GBLANOT01. TRNSFS-11

CC is expressed in dermatologic and gastrointestinal tissues,

CC especially those associated with inflammation and cell

CC proliferation. It shows homology to mouse N-acetylglicosamine
 CC 6-O-sulfotransferase. A polynucleotide comprising nucleotides
 CC 264-333 or 1272-1331 of the present sequence can be used as a DNA
 CC probe. The new human transferases and polynucleotides can be used
 CC in the diagnosis, prevention and treatment (including gene therapy
 CC and antisense therapy) of cancer, developmental disorders,
 CC gastrointestinal disorders, genetic disorders, immunological
 CC disorders, neurological disorders, reproductive disorders, and
 CC smooth muscle disorders.

XX

Sequence 2065 BP; 466 A; 575 C; 491 G; 533 T; 0 other:

Query Match 56.4%; Score 1153; DB 21; Length 2065;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 1593; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 106 CAAGGCTTCCACTTACGACACATGCTTACCTTAAAAAATGAAGCTCTGCTTTCT 165
 Db 152 CAAGGCTTCCACTTACGACACATGCTTACCTTAAAAAATGAAGCTCTGCTTTCT 211
 QY 166 GGTTCAGATGCGCATCTTGGCTTATCTTCCATGATGACGACCAACATCAGCTC 225
 Db 212 GGTTCAGATGCGCATCTTGGCTTATCTTCCATGATGACGACCAACATCAGCTC 271
 QY 226 CCTGCTATGAGGACACAGCCGACGACATGACGCTGCTGCTTCTTCTGCGCTC 285
 Db 272 CCTGCTATGAGGACACAGCCGACGACATGACGCTGCTGCTTCTTCTGCGCTC 331
 QY 286 TGGCTCTTCTTTTGGGGGACGCTTTTGGGACGACCCAGATGTTTCTTACCTGATGA 345
 Db 332 TGGCTCTTCTTTTGGGGGACGCTTTTGGGACGACCCAGATGTTTCTTACCTGATGA 391
 QY 346 GCGCGCGTGGACGCTGAGATGATCTTCAACACAGACGCGCGATGCTGACATGCG 405
 Db 392 GCGCGCGTGGACGCTGAGATGATCTTCAACACAGACGCGCGATGCTGACATGCG 451
 QY 406 TGTGCGGATCTGATACGCGCGCTTCTTGTGACATGACGCTTGTGATGCTACAT 465
 Db 452 TGTGCGGATCTGATACGCGCGCTTCTTGTGACATGACGCTTGTGATGCTACAT 511
 QY 466 GGAACCTGCTCCCGGACAGCTCAGCTTCTTCAAGTGGAGAAACGCGCGCTG 525
 Db 512 GGAACCTGCTCCCGGACAGCTCAGCTTCTTCAAGTGGAGAAACGCGCGCTG 571
 QY 526 TTTGCAACGCTGCTGATCAATCCCAAGATGAATATCCCGGCGCTCACTGAG 585
 Db 572 TTTGCAACGCTGCTGATCAATCCCAAGATGAATATCCCGGCGCTCACTGAG 630
 QY 586 GCTCCTGTGACATCAAGCCCTTTG-AGTGTGTGAGAAAGGCTGCGCTCTACAGCC 644
 Db 631 GCTCCTGTGACATCAAGCCCTTTGAGAGTTGTGGAGAAAGCCTGCGCTCTACAGCC 690
 QY 645 ACGTGTGCTCAGAGGCTGCTTCTTCAACCTGACAGTCCCTTACCCGCTGAAG 704
 Db 691 ACGTGTGCTCAGAGGCTGCTTCTTCAACCTGACAGTCCCTTACCCGCTGAAG 750
 QY 705 ACCGCTGCTCAACCTGATATGTCACGCTGCTGCGGACCGCGGCGCTGTTCCGTT 764
 Db 751 ACCGCTGCTCAACCTGATATGTCACGCTGCTGCGGACCGCGGCGCTGTTCCGTT 810
 QY 765 CCCGAGAACGCAAAAGGAGATCTCATGATTGACAGTGTGATGAGGAGCAGCATG 824
 Db 811 CCCGAGAACGCAAAAGGAGATCTCATGATTGACAGTGTGATGAGGAGCAGCATG 870
 QY 825 AGCAAAATCTCAGAGGAGCACCACCTTACTATGTGATGACAGTCTGCTCAAGCC 884
 Db 871 AGCAAAATCTCAGAGGAGCACCACCTTACTATGTGATGACAGTCTGCTCAAGCC 930
 QY 885 AGCTGGAGATCTCAAAACATTCAGTCTGCTGCGGACGCGCTCAGAGAGCTACCTGC 944
 Db 931 AGCTGGAGATCTCAAAACATTCAGTCTGCTGCGGACGCGCTCAGAGAGCTACCTGC 990
 QY 945 TTGTGCGCTATGAGACCTGCTGAGCCCTGTGCGCCAGACTTCCGGAATGTATGAAT 1004

```

Db      991 TTGTGCGCTATAGAGACCTGGCTCAGGCCCTGTGGCCAGACTTCCCAATGATGAT 1050
Oy      1005 TCGTGGGATTTGGAATTTCTTGGCCCATCTTCAAGCCCTGGGTGATATACATCACCCGAGGA 1064
Db      1051 TCGTGGGATTTGGAATTTCTTGGCCCATCTTCAAGCCCTGGGTGATATACATCACCCGAGGA 1110
Oy      1065 AGGGCATGGGTGACCAAGCTTTCACACAAATGCCAGGATGCCCTTAATGTCCTCCAGG 1124
Db      1111 AGGGCATGGGTGACCAAGCTTTCACACAAATGCCAGGATGCCCTTAATGTCCTCCAGG 1170
Oy      1125 CTGGGCGTGGTCTTGGCCCTATGAAAGTTTCTGATTTAGAAAGCTGTGGGATG 1184
Db      1171 CTGGGCGTGGTCTTGGCCCTATGAAAGTTTCTGATTTAGAAAGCTGTGGGATG 1230
Oy      1185 CCATGAATTTGCTGGGCTACGCCCAAGTCAGATCTGAAACAAAGAGAAACCTGTTC 1244
Db      1231 CCATGAATTTGCTGGGCTACGCCCAAGTCAGATCTGAAACAAAGAGAAACCTGTTC 1290
Oy      1245 TGGATCTTCTGCTACCTGAGCTGTCCCTGAGCAAAATCCACTAGAGGGTTGAGAGGCT 1304
Db      1291 TGGATCTTCTGCTACCTGAGCTGTCCCTGAGCAAAATCCACTAGAGGGTTGAGAGGCT 1350
Oy      1305 TTGCTGCACCTGATGTCAGCTCAGTCACTTCTGTAATGCTTGAAGCTTGGCTTAC 1364
Db      1351 TTGCTGCACCTGATGTCAGCTCAGTCACTTCTGTAATGCTTGAAGCTTGGCTTAC 1410
Oy      1365 ATCTGTGAGCCTTAATCATATCTGTGGGTATCAACACTGAGTGTGAGTGTGTCACAC 1424
Db      1411 ATCTGTGAGCCTTAATCATATCTGTGGGTATCAACACTGAGTGTGAGTGTGTCACAC 1470
Oy      1425 GTGCTCAACCAAGAGACTTTGTGTGCAATGCTGTGTGTAGAAAACAACATGAGGAAAC 1484
Db      1471 GTGCTCAACCAAGAGACTTTGTGTGCAATGCTGTGTGTAGAAAACAACATGAGGAAAC 1530
Oy      1485 TTATGTGAGCAGACATCCACCAAGTGAAGAGGATGTGCTTCTTCTTCTGATC 1544
Db      1531 TTATGTGAGCAGACATCCACCAAGTGAAGAGGATGTGCTTCTTCTTCTGATC 1590
Oy      1545 TTCTGTGTGGGAGACTTCAAGACTTTGTGGCTGAGGCTTATTAAGACAGACAG 1604
Db      1591 TTCTGTGTGGGAGACTTCAAGACTTTGTGGCTGAGGCTTATTAAGACAGACAG 1650
Oy      1605 TTATGATGGAATTTGATCCATAAAGCCTGTGCCAATCTGCTTGAATGGAATGATC 1664
Db      1651 TTATGATGGAATTTGATCCATAAAGCCTGTGCCAATCTGCTTGAATGGAATGATC 1710
Oy      1665 TTTCACCAAGAGCTCAGCAGATTTTCCACAGAGATGC 1703
Db      1711 TTTCACCAAGAGCTCAGCAGATTTTCCACAGAGATGC 1749

```

```

RESULT 5
AAK91803
ID      AAK91803 standard; cDNA; 877 BP.
AC      AAK91803;
XX      06-NOV-2001 (first entry)
DE      Human cDNA 5'-end sequence, SEQ ID NO: 263.
XX      Human; full length cDNA; cDNA synthesis; oligo-capping; sq.
XX      Homo sapiens.
XX      EP130094-A2.
XX      05-SEP-2001.
XX      07-JUL-2000; 2000EP-0114089.
XX      08-JUL-1999; 99JP-0194486.

```

```

PR      11-JAN-2000; 2000JP-0118774.
PR      02-MAY-2000; 2000JP-0183765.
XX      (HELI-) HELIX RES INST.
XX      Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI      Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX      WPI; 2001-524255/58.
PT      830 Primers useful for synthesizing full length cDNA clones and their
PT      use in genetic manipulation.
PS      Claim 2; SEQ ID NO 263; 1380bp + sequence listing; English.
XX      The invention relates to primers for synthesizing full length cDNA
XX      clones. 830 cDNA molecules encoding a human protein have been
XX      isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
XX      molecules have been determined. Primers for synthesizing the full length
XX      cDNA are useful for clarifying the function of the protein encoded by
XX      the cDNA. The full length clones were obtained by construction of full
XX      length enriched cDNA libraries that were synthesised by the oligo-capping
XX      method. The primers enable the production of the full length cDNA easily
XX      without any special methods. The present sequence is the nucleotide
XX      sequence of the 5'-end of a cDNA provided in the invention.
XX      Note: The sequence data for this patent did not form part of the printed
XX      specification, but was obtained in CD-ROM format directly from EPO.
SQ      Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;
Query Match      34.0%; Score 695; DB 22; Length 877;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy      106 CAAGTCTTCCACTTACAGACAAATGCTACTGCTTAAATAAATGAACTCTGCTTCTT 165
Db      47 CAAGTCTTCCACTTACAGACAAATGCTACTGCTTAAATAAATGAACTCTGCTTCTT 106
Oy      166 GGTTCCTCCAGATGGCATCTTGGCTATTTCTTCCACATGTACAGCCACAACTACGCTC 225
Db      107 GGTTCCTCCAGATGGCATCTTGGCTATTTCTTCCACATGTACAGCCACAACTACGCTC 166
Oy      107 GGTTCCTCCAGATGGCATCTTGGCTATTTCTTCCACATGTACAGCCACAACTACGCTC 166
Db      226 CCTGTATGAAGGACAGCCGAGGAGCAATGACAGTGTGCTTCTTCCGAGGCTC 285
Oy      167 CCTGTATGAAGGACAGCCGAGGAGCAATGACAGTGTGCTTCTTCCGAGGCTC 226
Db      226 CCTGTATGAAGGACAGCCGAGGAGCAATGACAGTGTGCTTCTTCCGAGGCTC 285
Oy      286 TGGCTCTTCTTGTGTGGGAGCTTTTGGGAGCAGCCAGATGTTTCTTACCTGATGA 345
Db      227 TGGCTCTTCTTGTGTGGGAGCTTTTGGGAGCAGCCAGATGTTTCTTACCTGATGA 286
Oy      346 GCCCGCTGGACAGTGTGATGACCTTCAAGCAGACAGCCGCTGATGCTGACATGGC 405
Db      287 GCCCGCTGGACAGTGTGATGACCTTCAAGCAGACAGCCGCTGATGCTGACATGGC 346
Oy      406 TGTGCGGATGATGATGCGGCGCTTCTTGTGCGCATGAGAGCTTTTGATGCTACAT 465
Db      347 TGTGCGGATGATGATGCGGCGCTTCTTGTGCGCATGAGAGCTTTTGATGCTACAT 406
Oy      466 GGAACCTGTGTCGCCGAGAGAGTCCAGCTCTTTCAGTGGAGAAACAGCCGGCCCTGTG 525
Db      407 GGAACCTGTGTCGCCGAGAGAGTCCAGCTCTTTCAGTGGAGAAACAGCCGGCCCTGTG 466
Oy      526 TTCTGCACTGCTGTGACATCATCCACAGATGAATATCATCCCGGGGCTACATGCA 585
Db      467 TTCTGCACTGCTGTGACATCATCCACAGATGAATATCATCCCGGGGCTACATGCA 526
Oy      586 GCTCCTGTGACAGCAACAGCCCTTGAAGTGTGAGGAGAGGCTGCGCTCTACAGCA 645
Db      527 GCTCCTGTGACAGCAACAGCCCTTGAAGTGTGAGGAGAGGCTGCGCTCTACAGCA 586
Oy      646 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTGAAGA 705
Db      587 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTGAAGA 646

```

QY 706 CCCCCTCCCTAACCTGATATCGTGACCTGTCGGGACCCCGGGCCGTTCCTCCGTC 765
 |||||||
 Db 647 CCCCCTCCCTAACCTGATATCGTGACCTGTCGGGACCCCGGGCCGTTCCTCCGTC 706
 |||||||
 QY 766 CCGAGAACGCACAAAGGAGATCTCATGATTGACA 800
 |||||||
 Db 707 CCGAGAACGCACAAAGGAGATCTCATGATTGACA 741
 |||||||
 RESULT 6
 AAK93921
 ID AAK93921 standard; cDNA: 877 BP.
 XX
 AC AAK93921;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human cDNA clone representative sequence, SEQ ID NO: 2381.
 XX
 KM Human, full length cDNA; cDNA synthesis; oligo-capping; ss.
 XX
 NS Homo sapiens.
 XX
 PN EP130094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-0114089.
 XX
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Oka T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI; 2001-52425/58.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 XX
 PS Example 11; SEQ ID NO 2381; 1380bp + sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence was used as the
 CC representative sequence from a human clone which was used in
 CC homology searches to identify the clone.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 CC
 XX
 SQ Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other:
 Query Match 34.0%; Score 695; DB 22; Length 877;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 106 CAAGGCTTCACCTTCAGACCAATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTCT 165
 |||||||
 Db 47 CAAGGCTTCACCTTCAGACCAATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTCT 106
 |||||||
 QY 166 GGTTCGCCAGATGGCCATCTTGCTATCTTCTTCACATGTACAGCCACAAATCAGCTC 225
 |||||||
 Db 107 GGTTCGCCAGATGGCCATCTTGCTATCTTCTTCACATGTACAGCCACAAATCAGCTC 166
 |||||||

QY 226 CCGTCTATGAAAGCAGACAGCCGAGCGATGACAGTGTGTTCTGTTCTTCTGCGCTC 285
 |||||||
 Db 167 CCGTCTATGAAAGCAGACAGCCGAGCGATGACAGTGTGTTCTGTTCTTCTGCGCTC 226
 |||||||
 QY 286 TGCGTCTCTTTTGTGGGGAGCTTTTGTGGGACAGCCAGCATTTTCTACTGATGGA 345
 |||||||
 Db 227 TGCGTCTCTTTTGTGGGGAGCTTTTGTGGGACAGCCAGCATTTTCTACTGATGGA 286
 |||||||
 QY 346 GCCCGCTGSCAGCTGTGATGACCTTCAGCAGACAGCCGCTGATGTCACATGGC 405
 |||||||
 Db 287 GCCCGCTGSCAGCTGTGATGACCTTCAGCAGACAGCCGCTGATGTCACATGGC 346
 |||||||
 QY 406 TGTCGGGATCTGATPAGGGGCGCTTCTTGTGGACATGAGCGTTTGTATGCTACAT 465
 |||||||
 Db 347 TGTCGGGATCTGATPAGGGGCGCTTCTTGTGGACATGAGCGTTTGTATGCTACAT 406
 |||||||
 QY 466 GGAACCTGTGTCGCCGAGACAGTCCAGCCTCTTCAGTGGAGAAACAGCCGCGCTGTG 525
 |||||||
 Db 407 GGAACCTGTGTCGCCGAGACAGTCCAGCCTCTTCAGTGGAGAAACAGCCGCGCTGTG 466
 |||||||
 QY 526 TTCTGACCTGCTGTGACATCATCCACAAGATGAATCATCCCGGGGCTCACTGAG 585
 |||||||
 Db 467 TTCTGACCTGCTGTGACATCATCCACAAGATGAATCATCCCGGGGCTCACTGAG 526
 |||||||
 QY 586 GCTCTGTGCACTCAACAGCCCTTGAAGTGTGGAGAGGCTGCGGCTCTACAGCA 645
 |||||||
 Db 527 GCTCTGTGCACTCAACAGCCCTTGAAGTGTGGAGAGGCTGCGGCTCTACAGCA 586
 |||||||
 QY 646 CGTGTGCTCAAGAGAGTGCGCTTCTTCACCTCAGTCCCTACCGGCTGTGAAAGA 705
 |||||||
 Db 587 CGTGTGCTCAAGAGAGTGCGCTTCTTCACCTCAGTCCCTACCGGCTGTGAAAGA 646
 |||||||
 QY 706 CCCCCTCCCTCACTGATATGTCGACCTGTCCGGAGCCCGGGCCGTTCCTGTC 765
 |||||||
 Db 647 CCCCCTCCCTCACTGATATGTCGACCTGTCCGGAGCCCGGGCCGTTCCTGTC 706
 |||||||
 QY 766 CCGAGAACGCACAAAGGAGATCTCATGATTGACA 800
 |||||||
 Db 707 CCGAGAACGCACAAAGGAGATCTCATGATTGACA 741
 |||||||
 RESULT 7
 ABV89280
 ID ABV89280 standard; cDNA: 505 BP.
 XX
 AC ABV89280;
 XX
 DT 13-DEC-2002 (first entry)
 XX
 DE Human colon cancer related cDNA SEQ ID NO 2595.
 XX
 KM Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
 KM ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200258534-A2.
 XX
 PD 01-AUG-2002.
 XX
 PF 19-NOV-2001; 2001WO-US43704.
 PF
 PR 20-NOV-2000; 2000US-252222P.
 PR 06-FEB-2001; 2001US-267011P.
 PR 28-MAR-2001; 2001US-279670P.
 PR 10-JUL-2001; 2001US-304037P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Stoik JA, Xu J, Chenaault RA, Meagher MJ, Secrist H, King GE;
 XX WPI; 2002-608400/65.

XX New isolated tumor colon polynucleotide and polypeptide, useful for the
 PT diagnosis, prevention and/or treatment of cancer, in particular colon
 PT cancer
 XX
 PS Claim 1: SEQ ID NO 2595; 266bp + Sequence Listing; English.
 XX
 CC The invention relates to a human colon tumour expressed polynucleotide
 CC (I) encoding a polypeptide (II), ABP67991-ABP67996) comprising: (i) any of
 CC 2600 fully defined nucleotide sequences (ABV6669-ABV8289); (ii) any of
 CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
 CC sequences that hybridize to (i), under moderately stringent conditions;
 CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
 CC degenerate variants of (i). The compositions and methods of the present
 CC invention are useful for the diagnosis, prevention and/or treatment of
 CC cancer, particularly colon cancer. (i) can be used in gene therapy and
 CC (i) and (ii) are useful in pharmaceutical compositions such as vaccines.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIFO
 CC at ftp.wipo.int/pub/published_pcr_sequences.
 CC
 SQ Sequence 505 BP; 120 A; 125 C; 119 G; 141 T; 0 other;
 XX
 Query Match 19.7%; Score 403; DB 24; Length 505;
 Best Local Similarity 100.0%; Pred. No. 2.6e-189; Mismatches 0; Gaps 0;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1301 GGCTTTGGTCCACCTGCTGTCAGCTTCTGTAATGCTTCTGAGCCTTC 1360
 DB 1 GGCTTTGGTCCACCTGCTGTCAGCTTCTGTAATGCTTCTGAGCCTTC 60
 QY 1361 CTACATCTGTGAGCCTTAACATGCTGTGGGTATCACTGAGTGTGATGTC 1420
 DB 61 CTACATCTGTGAGCCTTAACATGCTGTGGGTATCACTGAGTGTGATGTC 120
 QY 1421 ACACGTGCTCAAGCAGAAAGAGCTTTGTGCTATGCTTGTAGAAAAGACGAGG 1480
 DB 121 ACACGTGCTCAAGCAGAAAGAGCTTTGTGCTATGCTTGTAGAAAAGACGAGG 180
 QY 1481 AACCTTATGTGAGCAGACATCCACCACTGAAAGAGGTAATGCTTCTTCTTTT 1540
 DB 181 AACCTTATGTGAGCAGACATCCACCACTGAAAGAGGTAATGCTTCTTCTTTT 240
 QY 1541 GATCTTCTGCTGTGGGCACTTCAAGACTTTGTGGCTGAGGCTTATTAAGCAGAC 1600
 DB 241 GATCTTCTGCTGTGGGCACTTCAAGACTTTGTGGCTGAGGCTTATTAAGCAGAC 300
 QY 1601 ACAGTATCAGTGAATGATCATTAACCTCCCTGTCACATCTTGGCCCAATGGGGAATG 1660
 DB 301 ACAGTATCAGTGAATGATCATTAACCTCCCTGTCACATCTTGGCCCAATGGGGAATG 360
 QY 1661 GATCTTCTGCTGTGGGCACTTCAAGACTTTGTGGCTGAGGCTTATTAAGCAGAC 1703
 DB 361 GATCTTCTGCTGTGGGCACTTCAAGACTTTGTGGCTGAGGCTTATTAAGCAGAC 403
 XX
 RESULT 8
 ABR54724
 ID ABR54724 standard; cDNA; 517 BP.
 XX
 AC ABR54724;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Human colon cancer-associated cDNA, SEQ ID NO 194.
 XX
 KM Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200212280-A2.
 XX
 PD 14-FEB-2002.

XX
 PF 30-JUL-2001; 2001WO-US23826.
 XX
 PR 03-AUG-2000; 2000US-223265P.
 PR 02-OCT-2000; 2000US-237406P.
 PR 20-MAR-2001; 2001US-277495P.
 PR 03-JUL-2001; 2001US-302702P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Pyle RA, Xu J, Secretist H;
 XX
 DR WPI; 2002-257462/30.
 XX
 PT Novel polynucleotide encoding colon tumour polypeptides, useful as
 PT vaccines for treating colon cancers
 XX
 PS Claim 1; Page 206; 425bp; English.
 XX
 CC The invention relates to isolated polynucleotides (I) encoding colon
 CC tumour polypeptides (II). (I) is useful for stimulating an immune
 CC response in a patient and treating colon cancer in a patient.
 CC Oligonucleotides derived from (I) are useful for determining the presence
 CC of cancer in a patient. (I) and (II) are useful in pharmaceutical
 CC compositions, e.g. vaccines, and other compositions for the diagnosis
 CC and treatment of colon cancer. A composition comprising a first component
 CC selected from physiologically acceptable carriers and immunostimulants,
 CC and an antigen-presenting cell expressing (II) is useful for inhibiting
 CC development of cancer in a patient. (I) is useful in the design and
 CC preparation of ribozyme molecules for inhibiting expression of tumour
 CC polypeptides and (I). ABR54531-ABR55464 represent human colon cancer cDNA
 CC sequences of the invention.
 CC
 SQ Sequence 517 BP; 121 A; 130 C; 124 G; 142 T; 0 other;
 XX
 Query Match 19.7%; Score 403; DB 24; Length 517;
 Best Local Similarity 100.0%; Pred. No. 2.6e-189; Mismatches 0; Gaps 0;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1301 GGCTTTGGTCCACCTGCTGTCAGCTTCTGTAATGCTTCTGAGCCTTC 1360
 DB 13 GGCTTTGGTCCACCTGCTGTCAGCTTCTGTAATGCTTCTGAGCCTTC 72
 QY 1361 CTACATCTGTGAGCCTTAACATGCTGTGGGTATCACTGAGTGTGATGTC 1420
 DB 73 CTACATCTGTGAGCCTTAACATGCTGTGGGTATCACTGAGTGTGATGTC 132
 QY 1421 ACACGTGCTCAAGCAGAAAGAGCTTTGTGCTATGCTTGTAGAAAAGACGAGG 1480
 DB 133 ACACGTGCTCAAGCAGAAAGAGCTTTGTGCTATGCTTGTAGAAAAGACGAGG 192
 QY 1481 AACCTTATGTGAGCAGACATCCACCACTGAAAGAGGTAATGCTTCTTCTTTT 1540
 DB 193 AACCTTATGTGAGCAGACATCCACCACTGAAAGAGGTAATGCTTCTTCTTTT 252
 QY 1541 GATCTTCTGCTGTGGGCACTTCAAGACTTTGTGGCTGAGGCTTATTAAGCAGAC 1600
 DB 253 GATCTTCTGCTGTGGGCACTTCAAGACTTTGTGGCTGAGGCTTATTAAGCAGAC 312
 QY 1601 ACAGTATCAGTGAATGATCATTAACCTCCCTGTCACATCTTGGCCCAATGGGGAATG 1660
 DB 313 ACAGTATCAGTGAATGATCATTAACCTCCCTGTCACATCTTGGCCCAATGGGGAATG 372
 QY 1661 GATCTTCTGCTGTGGGCACTTCAAGACTTTGTGGCTGAGGCTTATTAAGCAGAC 1703
 DB 373 GATCTTCTGCTGTGGGCACTTCAAGACTTTGTGGCTGAGGCTTATTAAGCAGAC 415
 XX
 RESULT 9
 AAC76156
 ID AAC76156 standard; cDNA; 2988 BP.
 XX
 AC AAC76156;

XX 08-FEB-2001 (first entry)
 DT TTTGATGCTACATGGAACCTGTCCTCCGAGAGACAGTCCAGCTCTTCAAGTGGAGAAC 511
 XX 1480 TTTGATGCTACATGGAACCTGTCCTCCGAGAGACAGTCCAGCTCTTCAAGTGGAGAAC 1539
 DE Human OREF1711 polynucleotide sequence SEQ ID NO:3421.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antiparasitic; antiparkinsonian; noctropic; neuroprotective;
 KW anticonvulsant; osteopathic; antilarvial; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antihistaminic;
 KW antilivral; antibacterial; antifungal; antihemetic; antihydrolyd;
 KW antineuritic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 KW
 OS Homo sapiens.
 XX
 XX WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 XX (CURA-) CURAGEN CORP.
 PA
 PI Shinkets RA, Leach M;
 XX
 DR WPI; 2000-602362/57.
 DR P-PSDB; AAB41947.
 XX
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 2597-2599; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human OREF open reading frames 1 to 3161. The OREF
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiparasitic; antiparkinsonian; noctropic; neuroprotective;
 CC osteopathic; anticonvulsant; antilarvial; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antilivral; antibacterial; antifungal; antihemetic; antihydrolyd;
 CC antineuritic; and antineuritic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an OREF-associated disorder. The
 CC nucleic acids can be used to express OREF proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 XX Sequence 2988 BP; 613 A; 916 C; 842 G; 617 T; 0 other:
 SQ

QY 452 TTTGATGCTACATGGAACCTGTCCTCCGAGAGACAGTCCAGCTCTTCAAGTGGAGAAC 511
 DB 1480 TTTGATGCTACATGGAACCTGTCCTCCGAGAGACAGTCCAGCTCTTCAAGTGGAGAAC 1539
 QY 512 AGCGGGCCCTGTGTTCTGACACCTGCTGACATCATCCACAAGATGAATCATCCCC 571
 DB 1540 AGCGGGCCCTGTGTTCTGACACCTGCTGACATCATCCACAAGATGAATCATCCCC 1599
 QY 572 CGGGCTACTGAGGCTCTGTCAGTCAACAGCCCTTGAAGTGTGGAGAGGCTGCG 631
 DB 1600 CGGGCTACTGAGGCTCTGTCAGTCAACAGCCCTTGAAGTGTGGAGAGGCTGCG 1659
 QY 632 CGGCTCTACAGCAGCGTGTGCTCAAGAGGCGCTTCAACCTGACGTCCTGTCAC 691
 DB 1660 CGGCTCTACAGCAGCGTGTGCTCAAGAGGCGCTTCAACCTGACGTCCTGTCAC 1719
 QY 692 CGGCTCTGAAAGAGCCCTCCCTCAACCTGATATGTCAGCTGTCGAGGAGCCCGG 751
 DB 1720 CGGCTCTGAAAGAGCCCTCCCTCAACCTGATATGTCAGCTGTCGAGGAGCCCGG 1779
 QY 752 GCCGTG 757
 DB 1780 GCCGTG 1785
 RESULT 10
 ID ABR54794
 ID ABR54794 standard; cDNA; 389 BP.
 AC ABR54794;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Human colon cancer-associated cDNA, SEQ ID NO 264.
 KW Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.
 KW Homo sapiens.
 OS
 XX WO200212280-A2.
 XX
 PD 14-FEB-2002.
 XX
 PF 30-JUL-2001; 2001WO-US23826.
 XX
 PR 03-AUG-2000; 2000US-223265P.
 PR 02-OCT-2000; 2000US-237406P.
 PR 20-MAR-2001; 2001US-277495P.
 PR 03-JUL-2001; 2001US-302702P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Pyle RA, Xu J, Secretist R;
 XX
 DR WPI; 2002-257462/30.
 DR
 XX Novel polynucleotide encoding colon tumour polypeptides, useful as
 PT vaccines for treating colon cancers -
 PT
 PS Claim 1; Page 225; 425pp; English.
 PS
 CC The invention relates to isolated polynucleotides (I) encoding colon
 CC tumour polypeptides (II). (I) is useful for stimulating an immune
 CC response in a patient and treating colon cancer in a patient.
 CC Oligonucleotides derived from (I) are useful for determining the presence
 CC of cancer in a patient. (I) and (II) are useful in pharmaceutical
 CC compositions, e.g. vaccines, and other compositions comprising a first component
 CC and treatment of colon cancer. A composition comprising a first component
 CC selected from physiologically acceptable carriers and immunostimulants,
 CC and an antigen-presenting cell-expressing (II) is useful for inhibiting
 CC development of cancer in a patient. (I) is useful in the design and
 CC preparation of ribozyme molecules for inhibiting expression of tumour

CC polypeptides and (1). ABE54531-ABK5464 represent human colon cancer cDNA
 CC sequences of the invention.
 XX
 SQ Sequence 389 BP; 121 A; 83 C; 76 G; 109 T; 0 other;
 Query Match 5 6%; Score 114; DB 24; Length 389;
 Best Local Similarity 100.0%; Pred. No. 6.9e-46;
 Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1930 TGCCATCTGTAACTAAATCCCAATTAAGGTCGTGTAGATGCCCTTTTANG 1989
 DB 266 TGCCATCTGTAACTAAATCCCAATTAAGGTCGTGTAGATGCCCTTTTANG 325
 QY 1990 CTTCTTAATTATAGCAGTAAGTTCATTTTANGGATCCCTTAACAAAAA 2043
 DB 326 CTTCTTAATTATAGCAGTAAGTTCATTTTANGGATCCCTTAACAAAAA 379
 RESULT 11
 AAK92588/C
 ID AAK92588 standard; cDNA; 548 BP.
 XX
 AC AAK92588;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human cDNA 3'-end sequence, SEQ ID NO: 1048.
 XX
 KM Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP130094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-0114089.
 XX
 PR 08-JUL-1999; 99UP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 PI WPI; 2001-524255/58.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 XX
 PS Claim 3; SEQ ID NO 1048; 1380bp + sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is the nucleotide
 CC sequence of the 3'-end of a cDNA provided in the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 CC
 XX
 SQ Sequence 548 BP; 160 A; 109 C; 129 G; 148 T; 2 other;
 Query Match 5.1%; Score 105; DB 22; Length 548;
 Best Local Similarity 100.0%; Pred. No. 2e-41;
 Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1930 TGCCATCTGTAACTAAATCCCAATTAAGGTCGTGTAGATGCCCTTTTANG 1989
 DB 108 TGCCATCTGTAACTAAATCCCAATTAAGGTCGTGTAGATGCCCTTTTANG 49
 QY 1990 CTTCTTAATTATAGCAGTAAGTTCATTTTANGGATCCCTTAACAAAAA 2034
 DB 48 CTTCTTAATTATAGCAGTAAGTTCATTTTANGGATCCCTTAACAAAAA 4
 RESULT 12
 AAD24670
 ID AAD24670 standard; cDNA; 1647 BP.
 XX
 AC AAD24670;
 XX
 DT 12-MAR-2002 (first entry)
 XX
 DE Human drug metabolizing enzyme (DME)-5 cDNA.
 XX
 KM Human; drug metabolizing enzyme; gene therapy; autoimmune disorder;
 KM inflammatory disorder; acquired immune deficiency syndrome; infection;
 KM AIDS; adult respiratory distress syndrome; cell proliferative disorder;
 KM allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;
 KM cancer; endocrine disorder; hypothyroidism disorder; pituitary disorder;
 KM gastrointestinal disorder; metabolic disorder; developmental disorder;
 KM liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;
 KM goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
 KM DME-5; ss.
 XX
 KM Homo sapiens.
 XX
 OS
 XX
 FH Key Location/Qualifiers
 FH CDS 1..1188
 FT /tag- a
 FT /product- "Human DME-5 protein"
 FT sig-peptide 1..96
 FT /tag- b
 FT mat-peptide 97..1185
 FT /tag- c
 FT /product- "Human mature DME-5 protein #1"
 FT sig-peptide 1..105
 FT /tag- d
 FT mat-peptide 106..1185
 FT /tag- e
 FT /product- "Human mature DME-5 protein #2"
 XX
 PN WO200179468-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 12-APR-2001; 2001WO-US11869.
 XX
 PR 13-APR-2000; 2000US-197590P.
 PR 19-APR-2000; 2000US-198403P.
 PR 28-APR-2000; 2000US-200185P.
 PR 05-MAY-2000; 2000US-202234P.
 PR 11-MAY-2000; 2000US-203509P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Policky JL, Hafalia A, Burford N, Ring HZ, Lal P, Tribouley CM;
 PI Yao MG, Yue H, Tang YT, Patterson C, Das D, Sanjanwala MS;
 PI Gandhi AR, Reddy R, Khan FA, Baughn MR, Ramkumar J, Griffin JA;
 PI Au-Young J;
 XX
 DR WPI: 2002-066363/09.
 DR P-PSDB; AAE15438.
 XX
 PT Novel isolated human drug metabolizing enzymes referred as DME 1-10,
 PT useful for diagnosing, treating, or preventing disorders associated
 PT with aberrant expression of DME such as allergy, anemia, asthma,
 PT infertility -

PS Claim 5; Page 139; 143pp; English.

XX The invention relates to human drug metabolizing enzymes referred as
 CC DME and nucleic acid molecules encoding such enzymes. Polynucleotides
 CC of the invention are useful for assessing toxicity of test compounds
 CC and in gene therapy. Sequences of the invention are useful in the
 CC diagnosis, prevention and treatment of autoimmune/inflammatory
 CC disorders such as acquired immune deficiency syndrome (AIDS), adult
 CC respiratory distress syndrome, allergies, anaemia, atherosclerosis,
 CC asthma, autoimmune haemolytic anaemia, contact dermatitis, Crohn's
 CC disease, glomerulonephritis, Goodpasture's syndrome, gout, Grave's
 CC disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple
 CC sclerosis, osteoarthritis, osteoporosis, psoriasis, systemic lupus
 CC erythematosus, rheumatoid arthritis, scleroderma, ulcerative colitis,
 CC uveitis, viral, bacterial, fungal, parasitic, protozoal, helminthic
 CC infections; cell proliferative disorders such as actinic keratosis,
 CC arteriosclerosis, atherosclerosis, Duchenne and Becker dystrophy,
 CC hepatitis, Cushing's syndrome, cancers, myelodysplastic syndrome,
 CC epilepsy; endocrine disorders such as disorders of the hypothalamus
 CC and pituitary resulting from lesions such as primary brain tumours,
 CC adenomas, infarction associated with pregnancy, aneurysms, vascular
 CC malformations; eye disorders such as conjunctivitis, iritis, retinitis,
 CC glaucoma, pigmentosa, metabolic disorders such as Addison's disease,
 CC cystic fibrosis, diabetes, goitre, glycogen storage diseases,
 CC hypercholesterolaemia, hyperthyroidism, hypoglycaemia, lipid myopathies,
 CC Menkes syndrome, mannosidosis, obesity; gastrointestinal disorders such
 CC as dysphagia, gastric carcinoma, anorexia, nausea, gastroenteritis,
 CC hyperbilirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's
 CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental
 CC disorders. The present sequence is human DME-5 protein cDNA.

XX Sequence 1647 BP; 287 A; 560 C; 501 G; 299 T; 0 other;

XX

Query Match 2.9%; Score 59; DB 24; Length 1647;
 Best Local Similarity 100.0%; Pred. No. 1.4e-18;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCGTCCGCTCTACAGCCAGCGTGCTCAAGAGAGTGGCTTTCACCTGCAG 682
 |||||||
 Db 488 AGGCGTCCGCTCTCTACAGCCAGCGTGCTCAAGAGAGTGGCTTTCACCTGCAG 546

RESULT 13
 AAD02700 standard; cDNA; 1694 BP.

XX
 AC AAD02700;
 XX
 DT 02-MAY-2001 (first entry)

XX Human glycosyl sulfotransferase-4beta (GST-4beta) cDNA.

DE Human: glycosyl sulfotransferase-4beta; GST-4beta; Immunosuppressive;
 KW therapy; selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; rheumatic scleritis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 KW chromosome 16q23.1; ss.

XX
 OS Homo sapiens.

XX
 FH Key Location/Qualifiers
 FT 5'UTR 9..188
 FT CDS 189..1376
 FT /*tag- a
 FT /*tag- b
 FT /*product- "Human glycosyl sulfotransferase-4beta
 FT (GST-4beta)"
 FT /note- "CDS is referred as SEQ ID NO:21 in brief

FT description of the figures (page no: 4)"
 FT 1377..1694
 FT /*tag- c

XX 3'UTR
 XX
 XX WO200106015-A1.
 XX
 XX 25-JAN-2001.
 XX
 XX 19-JUL-2000; 2000MO-US19741.
 XX
 XX 20-JUL-1999; 99US-0144694.
 XX
 XX 13-JUL-2000; 2000US-0593828.
 XX
 XX (REGC) UNIV CALIFORNIA.
 XX
 XX Rosen SD, Lee JK, Hemmerich S;
 XX WPI; 2001-138471/14.
 XX P-PSDB: AAY72640.

DR New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 PT diagnostic and therapeutic agent screening applications

XX Claim 6; Fig 4A; 128pp; English.

XX The present sequence is human glycosyl sulfotransferase-4beta (GST-4
 CC beta) cDNA. GST-4 gene is found on chromosome 16q23.1.
 CC GST is a type 2 membrane protein useful for inhibiting a binding event
 CC between a selectin and a selectin ligand, which comprises contacting the
 CC selectin with a non-sulphated selectin ligand, GST and a small molecular
 CC agent that inhibits the sulphation activity of GST. GST is also useful
 CC in inhibiting a selectin mediated binding event. GST is useful in gene
 CC therapy to treat disorders such as acute or chronic inflammation,
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adenitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation.

XX Sequence 1694 BP; 259 A; 586 C; 514 G; 335 T; 0 other;

XX

Query Match 2.9%; Score 59; DB 22; Length 1694;
 Best Local Similarity 100.0%; Pred. No. 1.4e-18;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCGTCCGCTCTACAGCCAGCGTGCTCAAGAGAGTGGCTTTCACCTGCAG 682
 |||||||
 Db 676 AGGCGTCCGCTCTCTACAGCCAGCGTGCTCAAGAGAGTGGCTTTCACCTGCAG 734

RESULT 14
 AAD02699 standard; cDNA; 2044 BP.

XX
 AC AAD02699;
 XX
 DT 02-MAY-2001 (first entry)

XX Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #2.

DE Human: glycosyl sulfotransferase-4alpha; GST-4alpha; Immunosuppressive;
 KW therapy; selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; rheumatic scleritis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;

Fri Aug 15 08:09:31 2003

us-09-645-078-1_2.rng

Page 13

Db 834 AGGCTGCCGCTCCTACAGCCAGTGTGTCTCAAGAGAGTGCCTTCTTCAACCTGCAG 892

Search completed: August 14, 2003, 21:37:42
Job time : 545 secs

THIS PAGE BLANK (USTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 21:16:48 : Search time 3887 Seconds
(without alignments)
12774.368 Million cell updates/sec

Title: US-09-645-078-1
Perfect score: 2043
Sequence: 1 gaattccattgttgggta.....tgggattcctaaataaaaaa 2043

Scoring table: OLIGO_NTC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562182

Minimum DB seq length: 15

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estbta:*
2: em_esthum:*
3: em_estln:*
4: em_estlnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	685	33.5	954	12	BI823850 603039012
2	579	28.3	583	9	BSM082755 BK490456 Homo sapi
3	534	26.1	668	9	AL709927 DKF2p6860
4	479	23.4	571	12	BM129080 1f17c04.y

c	5	322	15.8	593	12	BM969292	BM969292	UI-CF-ENO
c	6	290	14.2	680	10	BE878439	BE878439	MRO-ET010
c	7	288	14.1	669	13	BU684395	BU684395	UI-CF-ENO
c	8	218	10.7	553	12	BM128370	BM128370	1f13d12.x
c	9	218	10.7	553	12	BM128831	BM128831	1f17c04.x
c	10	107	5.2	362	9	AM002418	AM002418	wu61903.x
c	11	105	5.1	417	9	AM572510	AM572510	xq18q11.x
c	12	101	4.9	358	10	BF056840	BF056840	7k10d12.x
c	13	74	3.6	419	9	AM572390	AM572390	xu09a06.x
c	14	59	2.9	525	10	BF197521	BF197521	7684a08.x
c	15	59	2.9	620	10	BF197521	BF197521	w146c01.x
c	16	59	2.9	695	10	BE857538	BE857538	7901a08.x
c	17	59	2.9	735	10	BE858652	BE858652	7901a09.x
c	18	55	2.7	721	28	AO741911	AO741911	H5-5569_B
c	19	47	2.3	436	9	AM081348	AM081348	xc41b05.x
c	20	27	1.3	507	12	BM246681	BM246681	K0741E04-
c	21	27	1.3	571	12	BM245312	BM245312	K0722H07-
c	22	27	1.3	695	12	BG964671	BG964671	602831875
c	23	27	1.3	783	12	BG963298	BG963298	602827716
c	24	27	1.3	852	12	BG966340	BG966340	602832826
c	25	27	1.3	1923	11	AK009113	AK009113	MUS mnsu
c	26	23	1.1	362	10	BE340900	BE340900	EST344988
c	27	23	1.1	397	9	AI345721	AI345721	LB84d04.x
c	28	23	1.1	687	14	CB167190	CB167190	GMM602702
c	29	23	1.1	692	14	BY711808	BY711808	BY711808
c	30	23	1.1	834	13	BO884877	BO884877	AGENCOURT
c	31	22	1.1	148	9	AI345073	AI345073	tb62b07.x
c	32	22	1.1	160	13	BO265877	BO265877	NISC_f09
c	33	22	1.1	167	9	AI253971	AI253971	qv34a05.x
c	34	22	1.1	167	9	AI733948	AI733948	qv34a05.x
c	35	22	1.1	171	9	AI246069	AI246069	qv34f03.x
c	36	22	1.1	187	9	AI251238	AI251238	qv35g05.x
c	37	22	1.1	187	9	AI792899	AI792899	qv35g05.y
c	38	22	1.1	191	9	AI251979	AI251979	qv37d09.x
c	39	22	1.1	196	9	AI792857	AI792857	qv34a05.y
c	40	22	1.1	253	9	AA076760	AA076760	7B01F06 C
c	41	22	1.1	290	9	AI345395	AI345395	lb82b03.x
c	42	22	1.1	361	9	AI311365	AI311365	ta47h05.x
c	43	22	1.1	493	10	BF869376	BF869376	1f3-ET011
c	44	22	1.1	501	28	BH287165	BH287165	CH230-113
c	45	22	1.1	516	9	AA689056	AA689056	v80c006.r

ALIGNMENTS

RESULT 1
BI823850 954 bp mRNA linear EST 04-OCT-2001
603039012P1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5179826 5',
LOCUS
DEFINITION
BI823850
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 954)
NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
Plate: LHAM1448 row: h column: 03
High quality sequence stop: 856.

FEATURES

Location/Qualifiers
1..954

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5179826"

/lab_host="DH10B"

/clone_11b="NIH_MGC_115"

/note="Organ: pooled brain, lung, testis; Vector:

PCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA

source anonymous pool of 6 male brains, age range 23-27; 1

male lung, age 27; and 1 male testis, age 69. Library is

oligo-dT primed and directionally cloned (EcoRV site is

destroyed upon cloning). Average insert size 1.8 kb.

Insert size range 1-3 kb. Library is normalized and

enriched for full-length clones and was constructed by C.

Gruber (Invitrogen). Research Genetics tracking code

021. Note: this is a NIH_MGC Library."

BASE COUNT

199 a 297 c 250 g 208 t

Query Match

Best Local Similarity 100.0%; Score 685; DB 12; Length 954;

Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

51 GCGCATGCGCGCGCTAGCAGTACGCTCTCAAAAGCAGGAGGCCCAAGCCACAGC 110
1 GCGCATGCGCGCGCTAGCAGTACGCTCTCAAAAGCAGGAGGCCCAAGCCACAGC 60
111 TCTTCACCTTCAGCAATGCTACTGCTCAAAATGAGGCTCCGCTTCTGCTT 170
61 TCTTCACCTTCAGCAATGCTACTGCTCAAAATGAGGCTCCGCTTCTGCTT 120
171 CCCAGATGCGCATCTTGGCTATCTTCCACATGATACGCCACAACATCAGCTCCCTGT 230
121 CCCAGATGCGCATCTTGGCTATCTTCCACATGATACGCCACAACATCAGCTCCCTGT 180
231 CATATAGGACAGCCCGGCGCATGACGCTGCTGCTTCTTCTTCTGCGCTGCGCT 290
181 CATATAGGACAGCCCGGCGCATGACGCTGCTGCTTCTTCTTCTGCGCTGCGCT 240
291 CTTCTTTTGTGGGCGAGCTTTTGGGCGACACCCAGATGTTTCTACCTGATGAGCCCG 350
241 CTTCTTTTGTGGGCGAGCTTTTGGGCGACACCCAGATGTTTCTACCTGATGAGCCCG 300
351 CCTGCGACGCTGTGATGACCTTCAAGCAGACGCGCTGATGCTGACATGCTGCTGC 410
301 CCTGCGACGCTGTGATGACCTTCAAGCAGACGCGCTGATGCTGACATGCTGCTGC 360
411 GGGATCTGATAGCGGCGCTCTTGTGTGGACATGAGCTTGTGATGCTACATGAGAC 470
361 GGGATCTGATAGCGGCGCTCTTGTGTGGACATGAGCTTGTGATGCTACATGAGAC 420
471 CTGATCCCGGAGACAGTCCAGCTCTTTCAGTGGAGAAACCGCGGCGCTGTGTTCTG 530
421 CTGATCCCGGAGACAGTCCAGCTCTTTCAGTGGAGAAACCGCGGCGCTGTGTTCTG 480
531 CACCTGCGCTGTGATCATCCACAAGATGAATATATCCCGGCGCTACTGAGGCTCC 590
481 CACCTGCGCTGTGATCATCCACAAGATGAATATATCCCGGCGCTACTGAGGCTCC 540
591 TGTGAGTCAACAGCCCTTGAAGTGTGAGAGAGCCCGGCTCCCTACGACGACGTGG 650
541 TGTGAGTCAACAGCCCTTGAAGTGTGAGAGAGCCCGGCTCCCTACGACGACGTGG 600
651 TGTGAGTCAACAGCCCTTGAAGTGTGAGAGAGCCCGGCTCCCTACGACGACGTGG 710
601 TGTGAGTCAACAGCCCTTGAAGTGTGAGAGAGCCCGGCTCCCTACGACGACGTGG 660
711 CCTCAACCTGCATATCGTGACCT 735
661 CCTCAACCTGCATATCGTGACCT 685

RESULT 2

HS082755 standard; RNA; EST; 583 BP.

AC BX490456;

SV BX490456.1

XX 09-MAY-2003 (Rel. 75, Created)

DT 09-MAY-2003 (Rel. 75, Last updated, Version 1)

XX Homo sapiens mRNA; EST DKFZp686G0680_r1 (from clone DKFZp686G0680)

DE EST; expressed sequence tag.

XX Homo sapiens (human)

XX Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;

OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.

XX [1]

RP 1-583

RA Ansoorge W., Krieger S., Regiert T., Rittmuller C., Schwager B.,

RA Mewes H.W., Well B., Amlid C., Osanger A., Robo G., Han M., Wiemann S.;

RT Submitted (07-MAY-2003) to the EMBL/GenBank/DBJ databases.

RL MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY

XX This is the 5' sequence of the clone insert.

CC Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

CC Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;

CC Heidelberg/Germany) within the CDNA sequencing consortium of

CC the German Genome Project.

CC No 51 sequence available.

CC This clone (DKFZp686G0680) is available at the RZPD in Berlin.

CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,

CC 14055 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC Key Location/Qualifiers

FH source

FT 1..583

FT /db_xref="taxon:9606"

FT /mol_type="mRNA"

FT /organism="Homo sapiens"

FT /clone="DKFZp686G0680"

FT /clone_11b="686 (synonym: hicc3). Vector pSport1_Sfi; host

FT DH10B; sites SfiIA + SfiIB"

FT /dev_stage="adult"

FT /tissue_type="CDNA-collection"

XX Sequence 583 BP; 120 A; 172 C; 154 G; 137 T; 0 other;

XX Query Match

Best Local Similarity 100.0%; Score 579; DB 2; Length 583;

Matches 579; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

41 GAGGAGAAAAAGCGCATGCGCGCGCTAGCAGTACGCTCTCAAAAGCAGGAGAAACCCA 100
1 GAGGAGAAAAAGCGCATGCGCGCGCTAGCAGTACGCTCTCAAAAGCAGGAGAAACCCA 60
101 AGCCACAAGGCTTCCACTTACGACCAATGCTACGCTCAAAAGGAGGAGGAGGAGG 160
61 AGCCACAAGGCTTCCACTTACGACCAATGCTACGCTCAAAAGGAGGAGGAGGAGG 120
161 TTTTGTGTTTCCAGATGCGCATCTTGTGCTATTTCTTCACATGATGACCAACATC 220
121 TTTTGTGTTTCCAGATGCGCATCTTGTGCTATTTCTTCACATGATGACCAACATC 180
221 AGCTCCCTGTATGAGGACAGCCCGGAGGCGCATGACGCTGCTGCTTCTGCTG 280
181 AGCTCCCTGTATGAGGACAGCCCGGAGGCGCATGACGCTGCTGCTGCTGCTGCTG 240

QY 281 CGCTCTGCTCTTCTTTTGTGGGACGCTTTTGGGACAGCCAGATGTTTCTACTG 340
 |||||||
 Db 241 CGCTCTGCTCTTCTTTTGTGGGACGCTTTTGGGACAGCCAGATGTTTCTACTG 300
 |||||||
 QY 341 ATGAGCCCGCTGCGACGCTGTGATACCTTCAAGCAGACACCGCTGTGCTGAC 400
 |||||||
 Db 301 ATGAGCCCGCTGCGACGCTGTGATACCTTCAAGCAGACACCGCTGTGCTGAC 360
 |||||||
 QY 401 ATGAGCTGCGGGATCTGATACGGGCGCTCTTGTGGGACATGACGCTTTGATGCC 460
 |||||||
 Db 361 ATGAGCTGCGGGATCTGATACGGGCGCTCTTGTGGGACATGACGCTTTGATGCC 420
 |||||||
 QY 461 TACATGACCTGTGCTCCCGGAGACAGTCCAGCCTTTTCAGTGGGAGAACGCGGACC 520
 |||||||
 Db 421 TACATGACCTGTGCTCCCGGAGACAGTCCAGCCTTTTCAGTGGGAGAACGCGGACC 480
 |||||||
 QY 521 CTGTGTTCTGACCTGCTGCTGTGACATATCCACAAATGATATCCCGGGCTCAC 580
 |||||||
 Db 481 CTGTGTTCTGACCTGCTGCTGTGACATATCCACAAATGATATCCCGGGCTCAC 540
 |||||||
 QY 581 TGCAGGCTCCTGTGCTGACATCAACAGCCCTTTGAGTGTG 619
 |||||||
 Db 541 TGCAGGCTCCTGTGCTGACATCAACAGCCCTTTGAGTGTG 579
 |||||||

RESULT 3
 AL709927 668 bp mRNA linear EST 12-JUN-2003
 LOCUS DKFP68602364_r1 686 (synonym: h1cc3) Homo sapiens cDNA clone
 DEFINITION DKFP68602364 5', mRNA sequence.
 AL709927
 AL709927.1 GI:19693282

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 688)
 Bahr A., Lauber J., Mewes H.W., Well B., Amlid C., Osanger A., Fobo
 G., Han M. and Wiemann S.
 EST (Bahr A., Lauber J., Mewes H.W., Well B., et al.)
 Unpublished
 CONTACT: Bahr A
 MIPs

INGOlstaedter Landstr.1, D-85764 Neuherberg, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by Olagen (Hilden/Germany) within the cDNA sequencing
 consortium of the German Genome Project.
 No sl sequence available.

FEATURES
 SOURCE
 location/Qualifiers
 1..668
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFP68602364"
 /tissue_type="human skeletal muscle"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="686 (synonym: h1cc3)"
 /note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
 cDNA-collection"

BASE COUNT 132 a 202 c 178 g 154 t 2 others
 ORIGIN

Query Match 26.1%; Score 534; DB 9; Length 668;
 Best Local Similarity 100.0%; Prid. NO. 2.3e-209;
 Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 AGAGAGAAAAAGCGCATGGCCCGGCTAGCAGTGAAGCTCTCAAAAGCAGGAGAGCC 99
 |||||||
 Db 6 AGAGAGAAAAAGCGCATGGCCCGGCTAGCAGTGAAGCTCTCAAAAGCAGGAGAGCC 65
 |||||||
 QY 100 AAGCACAAGGCTCTTCCACTTACAGACAAATGCTACTGCTTAAAAAATGAAGCTCTGCT 159
 |||||||
 Db 66 AAGCACAAGGCTCTTCCACTTACAGACAAATGCTACTGCTTAAAAAATGAAGCTCTGCT 125
 |||||||
 QY 160 GTTCTGTGTTCCCAAGATGGCAGTCTGCTCTTCTTCCACATGTACACCAACAT 219
 |||||||
 Db 126 GTTCTGTGTTCCCAAGATGGCAGTCTGCTCTTCTTCCACATGTACACCAACAT 185
 |||||||
 QY 220 CAGCTCCCTGCTATGAAGCAGACCCGAGCAGTCACTGCTGCTGCTCTGCTCTG 279
 |||||||
 Db 186 CAGCTCCCTGCTATGAAGCAGACCCGAGCAGTCACTGCTGCTGCTCTGCTCTG 245
 |||||||
 QY 280 GCGCTGTGCTCTTCTTGTGGGACGCTTTTGGGACCCAGATGTTTCTACT 339
 |||||||
 Db 246 GCGCTGTGCTCTTCTTGTGGGACGCTTTTGGGACCCAGATGTTTCTACT 305
 |||||||
 QY 340 GATGAGCCCGGCTGCGGACGCTGTGATGACCTTCAAGCAGACCGGCTGATGCTGCA 399
 |||||||
 Db 306 GATGAGCCCGGCTGCGGACGCTGTGATGACCTTCAAGCAGACCGGCTGATGCTGCA 365
 |||||||
 QY 400 CATGCTGTGCGGATGATACGCGCGCTCTTGTGACATGACGCTTTGATGC 459
 |||||||
 Db 366 CATGCTGTGCGGATGATACGCGCGCTCTTGTGACATGACGCTTTGATGC 425
 |||||||
 QY 460 CTACATGGAACCTGTGCTCCCGGAGACAGTCCAGCCTTTTCAGTGGGAGAACGCGG 519
 |||||||
 Db 426 CTACATGGAACCTGTGCTCCCGGAGACAGTCCAGCCTTTTCAGTGGGAGAACGCGG 485
 |||||||
 QY 520 CCTGTGTTCTGACCTGCTGCTGTGACATCAATCCCAAGATGAATATCCCGG 573
 |||||||
 Db 486 CCTGTGTTCTGACCTGCTGCTGTGACATCAATCCCAAGATGAATATCCCGG 539
 |||||||

RESULT 4
 BM129080 571 bp mRNA linear EST 12-MAR-2002
 LOCUS h17c04.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
 DEFINITION cDNA clone IMAGE:5676655 5', similar to TR:09Y3R3 09Y3R3
 N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE. ;, mRNA sequence.
 BM129080
 BM129080.1 GI:17123632

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 571)
 Melton D., Brown J., Kenty G., Permutt A., Lee C., Kaestner K.,
 Lemishka I., Scharer M., Brestelli J., Gradwohl G., Clifton S.,
 Hillier L., Marra M., Pape D., Wylie T., Martin J., Blisstein A.,
 Schmitt A., Theising B., Ritter E., Ronko I., Bennett J., Cardenas
 M., Gibbons M., McCann R., Cole R., Tsagarashvili R., Williams T.,
 Jackson X. and Bowers Y.
 Endocrine Pancreas Consortium
 Unpublished

OTHER ESTS: h17c04.x1
 CONTACT: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@lhop.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Juliana Brown
 (brown@fas.harvard.edu) This sequence now available from the IMAGE
 consortium, for clone orders contact: info@image.llnl.gov

High quality sequence stop: 434.

FEATURES
Location/Qualifiers
1. 571

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5676655"

/sex="Both"

/tissue_type="Islets of Langerhans"

/dev_stage="Adult"

/lab_host="DH10B"

/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"

/note="Organ: Pancreas; Vector: pSPORL1; Site: 1; Not 1; Site: 2; Sal 1; Starting library constructed using Superscript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an Ecot of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 109 a 169 c 143 g 150 t

Query Match 23 4%; Score 479; DB 12; Length 571;

Best Local Similarity 100.0%; Pred. No. 1.2e-186; Mismatches 4/9; Conservative 0; Indels 0; Gaps 0;

106 CAGGCTCTTCACATTCAGCACAATGCTACTGCTTAAATAAGACCTCTGCTTTCT 165
133 CAGGCTCTTCACATTCAGCACAATGCTACTGCTTAAATAAGACCTCTGCTTTCT 132
166 GGTTCCTCCAGATGGCATTTGGCTTATTTCCACATGTACAGCCACAATCAGCTC 225
133 GGTTCCTCCAGATGGCATTTGGCTTATTTCCACATGTACAGCCACAATCAGCTC 192
226 CCGTCTATGAAGGACACCCGAGGAGCATGCTGCTGCTTCTTCCGGGCGCTC 285
193 CCGTCTATGAAGGACACCCGAGGAGCATGCTGCTGCTTCTTCCGGGCGCTC 252
286 TGGCTCTCTTTTGTGGGAGCCTTTTGGGAGCACCAGATGTTTCTACCTGATGA 345
253 TGGCTCTCTTTTGTGGGAGCCTTTTGGGAGCACCAGATGTTTCTACCTGATGA 312
346 GCGCGCTGGGACGCTGATGACCTTCAAGCAGACACCGCTGATGCTGACATGGC 405
313 GCGCGCTGGGACGCTGATGACCTTCAAGCAGACACCGCTGATGCTGACATGGC 372
406 TGTGGGAGATGTATACGGGCGCTTCTTGTGCGACATGAGGCTTTGATGCTACAT 465
373 TGTGGGAGATGTATACGGGCGCTTCTTGTGCGACATGAGGCTTTGATGCTACAT 432
466 GGAACCTGTCTCCCGGAGACAGTCCAGCTTTTCAAGTGGGAGAACAGCCGGCCCTGTG 525
433 GGAACCTGTCTCCCGGAGACAGTCCAGCTTTTCAAGTGGGAGAACAGCCGGCCCTGTG 492
526 TTTCTGACCTGCTGTGACATCATCCCAAGATGAATCATCCCCGGGGCTACATGCA 584
493 TTTCTGACCTGCTGTGACATCATCCCAAGATGAATCATCCCCGGGGCTACATGCA 551

RESULT 5
LOCUS BM969292/ 593 bp mRNA linear EST 20-FEB-2003
DEFINITION UI-CF-ENO-acp-1-21-0-UI.s1 UI-CF-ENO Homo sapiens cDNA clone
ACCESSION BM969292
VERSION BM969292.1 GI:19586879

KEYWORDS
SOURCE
ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL
MEDLINE
PUBMED

COMMENT

EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 593)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889348
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA=yes.

FEATURES
source

1. 593
/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-CF-ENO-acp-1-21-0-UI"

/tissue_type="Primary Lung Cystic Fibrosis Epithelial Cells"

/dev_stage="Adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone_lib="UI-CF-ENO"

/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site: 1: EcoR I; Site: 2: Not I; UI-CF-ENO is a cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGCT.

TAG LIB-UI-CF-ENO

TAG_TISSUE=Human Lung Epithelial Cell Lines untreated LPS

6hr to LPS 24h

TAG_SEQ=CTGCTCAGCT

BASE COUNT 157 a 127 c 137 g 172 t

Query Match 15 8%; Score 322; DB 12; Length 593;

Best Local Similarity 100.0%; Pred. No. 6.6e-122; Mismatches 3/2; Conservative 0; Indels 0; Gaps 0;

1382 ACATGCTGTGGGATACACATGAGTGTGTTGTCACACAGTCTCAACAGAGA 1441
1442 CTTTGTGTCCATGCTTGTCTAGAAAACAGACTGGGAACTTATGTGACAGACAT 1501
533 CTTTGTGTCCATGCTTGTCTAGAAAACAGACTGGGAACTTATGTGACAGACAT 474
1502 CCCACAGTGAACAGAGGATATGCTCTTTCTTTTCTTGATCTTCTGCGGACAGAC 1561


```

|||||
Db 473 CCCACAGTGAAGAGGGGATGCTCTTCTTTCTTGATCTCTGTCGGGACAGC 414
Oy 1562 TTGAGAGCTTTGGGCGCTGGAGGCTTTAAGCAGCAGCATGTGCGAATGATC 1621
Db 413 TTTCAGAGCTTTGGGCGCTGGAGGCTTTAAGCAGCAGCATGTGCGAATGATC 354
Oy 1622 CATAAAGCTCCCTGTCACATCTTGCCCAATGGGAAATGATCTTTCACCAAGAGCTCA 1681
Db 353 CATAAAGCTCCCTGTCACATCTTGCCCAATGGGAAATGATCTTTCACCAAGAGCTCA 294
Oy 1682 CCAGCATTTTCACAGAGATGC 1703
Db 293 CCAGCATTTTCACAGAGATGC 272

RESULT 6
BF878439 680 bp mRNA linear EST 17-JAN-2001
LOCUS MRO-ET0109-191100-002-h06 ET0109 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF878439
VERSION BF878439.1 GI:12268569
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 680)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=MR0&t2=MR0-ET0109-
191100-002-h06&t3=2000-11-19&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 641.
Location/Qualifiers
1. 680
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_1lb="ET0109"
/note="Organ: Lung,tumor; Vector: puc18; Site:1: SmaI;
Site:2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research).
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 168 a 164 c 163 g 184 t 1 others
ORIGIN
Query Match 14.2%; Score 290; DB 10; Length 680;
Best Local Similarity 100.0%; Pred. No. 9.9e-109;

```

```

Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1414 TGTGTCACAGCTGTGTCAGAGAGAGCTTTGTGTCAGTCTGTAGAAACAG 1473
Db 106 TGTGTCACAGCTGTGTCAGAGAGAGCTTTGTGTCAGTCTGTAGAAACAG 165
Oy 1474 ACTGGGAACCTTTAAGTAGAGAGAGCAGCAGCAGCAGTAAACAGGATATGCTCTTC 1533
Db 166 ACTGGGAACCTTTAAGTAGAGAGAGCAGCAGCAGCAGTAAACAGGATATGCTCTTC 225
Oy 1534 TTTCTGTATCTTCTGTCGTGGGAGAGCTTCAGAGACTTTGTGGCTGGAGGCTATTAA 1593
Db 226 TTTCTGTATCTTCTGTCGTGGGAGAGCTTCAGAGACTTTGTGGCTGGAGGCTATTAA 285
Oy 1594 GCACGACAGATATGAGTGAATGATATGATACCTGCTGCACATCTGCCCATG 1653
Db 286 GCACGACAGATATGAGTGAATGATATGATACCTGCTGCACATCTGCCCATG 345
Oy 1654 GGAATGATCTTTTCACCAAGAGCTCAGCAGCATTTTCCACAGAGATGC 1703
Db 346 GGAATGATCTTTTCACCAAGAGCTCAGCAGCATTTTCCACAGAGATGC 395

RESULT 7
B0684395/c 669 bp mRNA linear EST 07-OCT-2002
LOCUS UI-CF-ENO-aco-f-08-0-UI.s1 UI-CF-ENO Homo sapiens cDNA clone
DEFINITION UI-CF-ENO-aco-f-08-0-UI 3', mRNA sequence.
ACCESSION B0684395
VERSION B0684395.1 GI:23537302
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 669)
Bonaldi,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue procurement: Dr. M. J. Welsh, University of Iowa
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLY-A-Tes.
Location/Qualifiers
1. 669
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_1lb="UI-CF-ENO"
/note="Organ: Lung; Vector: pRT73-Pac (Pharmacia) with a
modified polylinker; Site:1: EcoR I; Site:2: Not I;
UI-CF-ENO is a cDNA library containing the following
tissue(s): Primary lung Cystic Fibrosis Epithelial Cells."
FEATURES
source

```

The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)₁₈ tail. The sequence tag for this library is CTGCTCAGGT.

TAG LIB-UI-CF-ENO
TAG-TISSUE-Human Lung Epithelial Cell Lines untreated LPS

6hr to LPS 24h
TAG-SEQ-CTGCTCAGGT

BASE COUNT 191 a 138 c 147 g 193 t

Query Match 14.1%; Score 288; DB 13; Length 669;
Best Local Similarity 100.0%; Pred. No. 6.7e-106; Mismatches 0; Indels 0; Gaps 0;
Matches 288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1416 TGTCCACAGCTGCTCAGCAGAGAGACTTTGTGTCATGCTGTGCTAGAAACAGAC 1475
|||||
DB 636 TGTCCACAGCTGCTCAGCAGAGAGACTTTGTGTCATGCTGTGCTAGAAACAGAC 577
|||||
QY 1476 TGGGAACTTATGTAGACAGACATCCACAGAGGTATTGCTCTTCTTCTT 1535
|||||
DB 576 TGGGAACTTATGTAGAGACAGACATCCACAGAGGTATTGCTCTTCTTCTT 517
|||||
QY 1536 TTCTTGATCTTCCTGCTGCGGAGACTTCAGACATTTGGCTGGAGCCATTAAAG 1595
|||||
DB 516 TTCTTGATCTTCCTGCTGCGGAGACTTCAGACATTTGGCTGGAGCCATTAAAG 457
|||||
QY 1596 AGCAGACAGTATGATGTAATTCATTAACCTCCGTGCCACATCTTCCCAATGG 1655
|||||
DB 456 AGCAGACAGTATGATGTAATTCATTAACCTCCGTGCCACATCTTCCCAATGG 397
|||||
QY 1656 GAATGATCTTCACCAAGAGCTCACGACATTTCCACAGAGATGC 1703
|||||
DB 396 GAATGATCTTCACCAAGAGCTCACGACATTTCCACAGAGATGC 349
|||||

RESULT 8
BM128370/c 553 bp mRNA linear EST 12-MAR-2002
LOCUS BM128370
DEFINITION lfi3d12.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
ACCESSION CDNA clone IMAGE:5676335 3', mRNA sequence.
VERSION BM128370
KEYWORDS BM128370.1 GI:17122922
SOURCE EST.
ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 553)

TITLE
JOURNAL
COMMENT
Unpublished
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on

obtaining a clone please contact: Juliana Brown
(brownjefas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
High quality sequence stop: 425.
Location/Qualifiers

FEATURES
source

1..553
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5676335"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/note="Organ: Pancreas; Vector: pSPORT1; Site: 1; Not 1;
Site: 2; Sal 1; Starting library constructed using
Superscript Plasmid Library kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 microgram PCR product
representing library inserts and hybridized to an EcoT of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."

BASE COUNT 163 a 111 c 127 g 152 t

Query Match 10.7%; Score 218; DB 12; Length 553;
Best Local Similarity 100.0%; Pred. No. 5.3e-79; Mismatches 0; Indels 0; Gaps 0;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1486 TATGAGAGACATCCACAGCAGTGAACAGGTATGCTCTCTCTTCTTCTTATCT 1545
|||||
DB 553 TATGAGAGACATCCACAGCAGTGAACAGGTATGCTCTCTCTTCTTCTTATCT 494
|||||
QY 1546 TCCTGTCTGGGAGACTTCAGAGACTTGTGCGCTGAGAGCCATTAAACAGACAGT 1605
|||||
DB 493 TCCTGTCTGGGAGACTTCAGAGACTTGTGCGCTGAGAGCCATTAAACAGACAGT 434
|||||
QY 1606 ATCAGTGAATTCATTAACCTCCGTGCCACATCTTCCCAATGGGAATGATCT 1665
|||||
DB 433 ATCAGTGAATTCATTAACCTCCGTGCCACATCTTCCCAATGGGAATGATCT 374
|||||
QY 1666 TTCACCAAGAGCTCACGACATTTCCACAGAGATGC 1703
|||||
DB 373 TTCACCAAGAGCTCACGACATTTCCACAGAGATGC 336
|||||

RESULT 9
BM128831/c 553 bp mRNA linear EST 12-MAR-2002
LOCUS BM128831
DEFINITION lfi7c04.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
ACCESSION CDNA clone IMAGE:5676655 3', mRNA sequence.
VERSION BM128831
KEYWORDS BM128831.1 GI:17123383
SOURCE EST.
ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 553)

TITLE
JOURNAL
COMMENT
Unpublished
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on

JOURNAL
COMMENT

Unpublished
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
High quality sequence, stop: 425.

FEATURES
SOURCE

```

BASE COUNT      163 a      111 c      127 g      152 t
ORIGIN
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_image="5676655"
/sex="both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_id="Melton Normalized Human Islet 4 N4-His 1"
/note="Organ: Pancreas; Vector: pSPOR1; Site 1: Not 1; Site 2: Sal 1; Starting library constructed using Superscript plasmid library kit (Life Technologies). cDNA made by oligo-dT priming. Size selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an Eco of 20 single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

```

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
1 (bases 1 to 362)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer genome Anatomy Project (CGAP),
tumor gene index

JOURN
COMMENT

FEATURES
source
 Email: ccgabs-r@mail.nih.gov
 Tissue procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatma
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.dio.llnl.gov/bdrrp/image/image.html
 Seq primer: -400p from GIBCO.
 Location/Qualifiers
 I . 362

BASE COUNT
ORIGIN

Query Match	5.2%	Score 107;	DB 9;	Length 362;
Best Local Similarity	100.0%	Pred. No. 3	7e-33;	
Matches 107;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1930	TGCCATCTGTTAACTACTAAATTCACAAATAGGCTCTGTTAGAAATGCCCTTTTATG	19899	
DB	107	TGCCATCTGTTAACTACTAAATTCACAAATAGGCTCTGTTAGAAATGCCCTTTTATG	48	
QY	1990	CTTCTTAATTAATAGCAGTAAGTTCATTTTGTGGAACTCTAAA	2036	
DB	47	CTTCTTAATTAATAGCAGTAAGTTCATTTTGTGGAACTCTAAA	1	

**JOURN
COMMENT**

Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/resources.shtml
 Seq primer: -40UP from gibco
 High quality sequence stop: 416.
 Location/Qualifiers
 1. 417

FEATURES
 source
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2751044"
 /tissue_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"
 /lab_host="DH10B"
 /clone_1lib="NCI-CGAP-Utl1"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"
 BASE COUNT
 140 a 65 c 80 g 132 t
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Pred. No. 2.3e-32;
 Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1930 TCCCATCTGTATTAACATAATCCCAATTAAGCTCTGTTAGATGCCCTTTTATG 1989
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 266 TCCCATCTGTATTAACATAATCCCAATTAAGCTCTGTTAGATGCCCTTTTATG 207
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 1990 CTCTTAATTATTAGACAGTAATAGTTCATTTTATGAGATCCAA 2034
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 206 CTCTTAATTATTAGACAGTAATAGTTCATTTTATGAGATCCAA 162
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 12
 BF056840/c 358 bp mRNA 1linear EST 16-OCT-2000
 LOCUS 7k10d12.x1 NCL_CGAP_GC6 Homo sapiens CDNA clone IMAGE:3443783 3,
 DEFINITION mRNA sequence.
 ACCESSION BF056840
 VERSION BF056840.1 GI:10810736
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 358)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapsb-remail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
 DNA Sequencing Arrayed by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: inf@image.llnl.gov
 Seq primer: -40UP from gibco.

FEATURES
 source
 /organism="Homo sapiens"
 1. 358
 Location/Qualifiers
 /mol_type="mRNA"

/db_xref="taxon:9606"
 /clone="IMAGE:3443783"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /clone_1lib="NCI-CGAP GC6"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP-GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1257086-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT
 107 a 68 c 82 g 101 t
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Pred. No. 1.1e-30;
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1932 CCATCTGTTAATACATAATCCCAATTAAGCTCTGTTAGATGCCCTTTTATGCT 1991
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 101 CCATCTGTTAATACATAATCCCAATTAAGCTCTGTTAGATGCCCTTTTATGCT 42
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 1992 TCTTAATTATTAGACAGTAATAGTTCATTTTATGAGATCCT 2032
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 41 TCTTAATTATTAGACAGTAATAGTTCATTTTATGAGATCCT 1

RESULT 13
 AM572390/c 419 bp mRNA 1linear EST 13-MAR-2000
 LOCUS XU09a06.x2 NCL_CGAP_C014 Homo sapiens CDNA clone IMAGE:2799634 3,
 DEFINITION mRNA sequence.
 ACCESSION AM572390
 VERSION AM572390.1 GI:7237123
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 419)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapsb-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing Arrayed by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/resources.shtml
 Seq primer: -40UP from gibco
 High quality sequence stop: 343.
 Location/Qualifiers
 1. 419

FEATURES
 source
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2799634"
 /tissue_type="moderately-differentiated adenocarcinoma"
 /lab_host="DH10B"
 /clone_1lib="NCI-CGAP_C014"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.7 kb. Life Technologies catalog #: 11531-019"

BASE COUNT
 122 a 78 c 89 g 130 t

ORIGIN

Query Match 3.6%; Score 74; DB 9; Length 419;

Best Local Similarity 100.0%; Pred. No. 1.4e-19;

Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1630 TCCCTGCCACATCTTCCCAATGGGAGATGATCTTTCACCAAGAGCTCACCAGATT.1689
 |||||||
 DB 419 TCCTGTCCACATCTTCCCAATGGGAGATGATCTTTCACCAAGAGCTCACCAGATT 360
 |||||||
 QY 1690 TTCACACAGATGC 1703
 |||||||
 DB 359 TTCACACAGATGC 346

RESULT 14

BF197521/c

LOCUS

DEFINITION 7684608.x1 NCI-CGAP_Kid11 Homo sapiens CDNA clone IMAGE:3642903 3'
 similar to TR:Q9UBX3 Q9UBX3 N-ACETYLGLUCOSAMINE
 6-O-SULFOTRANSFERASE. ; mRNA sequence.

ACCESSION

BF197521

VERSION

BF197521.1

ORGANISM

Homo sapiens (human)

REFERENCE

1 (bases 1 to 525)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

Authors

NCI-CGAP

Journal

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL, send email to:

Info@image.llnl.gov

High quality sequence stop: 451.

Location/Qualifiers

1..525

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:3642903"

/lab_host="DH10B"

/clone_1lb="NCI-CGAP_Kid11"

/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site1: Not I; Site2: Eco RI;

Plasmid DNA from the normalized library NCI-CGAP_Kid3

prepared, and ss circles were made in vitro. Following HAP

hybridization, this DNA was used as tracer in a subtracting

from a pool of 5,000 clones made from the same library

(cloneIDs 1322376-1323911, 1456007-1456775, and

1500352-1502855). Subtraction by Bento Soares and M.

Fatima Bonaldo.

BASE COUNT 90 a 163 c 187 g 85 t

ORIGIN

Query Match 2.9%; Score 59; DB 10; Length 525;
 Best Local Similarity 100.0%; Pred. No. 2e-13;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 15

A1824100/c

LOCUS

DEFINITION

A1824100.1 NCI-CGAP_Lu19 Homo sapiens CDNA clone IMAGE:2405856 3'

similar to TR:Q75667 Q75667 DU7116.4 ; contains PFR5

repetitive element ; mRNA sequence.

A1824100

ACCESSION

A1824100.1

VERSION

A1824100.1

ORGANISM

Homo sapiens (human)

REFERENCE

1 (bases 1 to 620)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

Authors

NCI-CGAP

Journal

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

www.bio.llnl.gov/dbp/image/image.html

Insert Length: 917 Std Error: 0.00

Seq primer: -400P from Glbco

High quality sequence stop: 490.

Location/Qualifiers

1..620

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2405856"

/tissue="squamous cell carcinoma, poorly

differentiated (4 pooled tumors, including primary and

metastatic)"

/dev_stage="adult"

/lab_host="DH10B (phage-resistant)"

/clone_1lb="NCI-CGAP_Lu19"

/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a

modified polylinker; 1st strand cDNA was prepared from

pooled lung tumor tissue, and was then primed with a Not I

- oligo(dT) primer. Double-stranded cDNA was ligated to

Eco RI adaptors (Pharmacia), digested with Not I and

cloned into the Not I and Eco RI sites of the modified

pT73 vector. Library went through one round of

normalization. Library constructed by Bento Soares and M.

Fatima Bonaldo.

BASE COUNT 106 a 195 c 227 g 89 t 3 others

ORIGIN

Query Match 2.9%; Score 59; DB 9; Length 620;
 Best Local Similarity 100.0%; Pred. No. 1.9e-13;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: August 15, 2003, 00:44:17

Job time: 3888 secs

QY 624 AGGCGTGGCGCTCTACAGCAGCGTGTGCTCAAGAGAGTGGCTTCTTCAACTGCAG 682
 |||||||
 DB 314 AGGCGTGGCGCTCTACAGCAGCGTGTGCTCAAGAGAGTGGCTTCTTCAACTGCAG 256

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 21:19:32 : Search time 142 Seconds
(without alignments)
6350.322 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043
Sequence: 1 gaattccattgtgttggtga.....tgggattccttaaaaaaaaaa 2043

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1092396

Minimum DB seq length: 15

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_MN:*
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1547	75.7	2032	US-09-045-284A-1	Sequence 1, Appl
2	1547	75.7	2032	US-09-190-911-2	Sequence 2, Appl
3	1153	56.4	2065	US-09-786-240-26	Sequence 26, Appl
4	37	1.8	37	US-09-045-284A-7	Sequence 7, Appl
5	37	1.8	37	US-09-190-911-7	Sequence 7, Appl
6	19	0.9	3792	US-08-992-334-1	Sequence 1, Appl
7	19	0.9	3792	US-08-302-752-1	Sequence 1, Appl
8	19	0.9	5234	US-08-992-334-2	Sequence 1, Appl
9	19	0.9	5234	US-08-302-752-2	Sequence 1, Appl
10	19	0.9	6722	US-08-992-334-3	Sequence 2, Appl
11	19	0.9	6722	US-08-302-752-3	Sequence 2, Appl
12	18	0.9	23	US-08-778-494B-112	Sequence 3, Appl
13	18	0.9	47	US-09-045-284A-8	Sequence 8, Appl
14	18	0.9	47	US-09-190-911-8	Sequence 8, Appl
15	18	0.9	150	US-09-439-313-320	Sequence 320, App
16	18	0.9	150	US-09-352-616A-320	Sequence 320, App
17	18	0.9	150	US-09-232-149A-320	Sequence 320, App
18	18	0.9	201	US-09-020-956-78	Sequence 78, Appl
19	18	0.9	201	US-09-030-607-78	Sequence 78, Appl
20	18	0.9	201	US-09-439-313-78	Sequence 78, Appl
21	18	0.9	201	US-09-352-616A-78	Sequence 78, Appl
22	18	0.9	201	US-09-232-149A-78	Sequence 78, Appl
23	18	0.9	202	US-09-030-607-180	Sequence 180, App
24	18	0.9	202	US-09-439-313-180	Sequence 180, App
25	18	0.9	202	US-09-352-616A-180	Sequence 180, App
26	18	0.9	202	US-09-232-149A-180	Sequence 180, App
27	18	0.9	536	US-09-020-956-69	Sequence 69, Appl

28	18	0.9	536	US-09-030-607-69	Sequence 69, Appl
29	18	0.9	536	US-09-439-313-69	Sequence 69, Appl
30	18	0.9	536	US-09-352-616A-69	Sequence 69, Appl
31	18	0.9	536	US-09-232-149A-69	Sequence 69, Appl
32	18	0.9	699	US-09-107-532A-1333	Sequence 1333, Ap
33	18	0.9	822	US-09-020-956-29	Sequence 29, Appl
34	18	0.9	822	US-09-030-607-29	Sequence 29, Appl
35	18	0.9	822	US-09-439-313-29	Sequence 29, Appl
36	18	0.9	822	US-09-352-616A-29	Sequence 29, Appl
37	18	0.9	822	US-09-232-149A-29	Sequence 29, Appl
38	18	0.9	1197	US-09-257-179-26	Sequence 26, Appl
39	18	0.9	1368	US-09-118-554-61	Sequence 61, Appl
40	18	0.9	1368	US-09-118-627-61	Sequence 61, Appl
41	18	0.9	1368	US-09-602-877A-61	Sequence 61, Appl
42	18	0.9	1913	US-09-599-360B-40	Sequence 40, Appl
43	18	0.9	1927	US-09-336-356-66	Sequence 66, Appl
44	18	0.9	4351	US-09-634-238-25	Sequence 25, Appl
45	17	0.8	141	US-08-036-555B-161	Sequence 161, App

ALIGNMENTS

RESULT 1					
US-09-045-284A-1					
: Sequence 1, Application US/09045284A					
: Patent No. 6265192					
GENERAL INFORMATION:					
: APPLICANT: Bistup, Annette					
: APPLICANT: Rosen, Steven D.					
: APPLICANT: Hemmerich, Stefan					
: TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3					
: FILE REFERENCE: 6510-107U51					
: CURRENT APPLICATION NUMBER: US/09/045, 284A					
: CURRENT FILING DATE: 1998-03-20					
: NUMBER OF SEQ ID NOS: 9					
: SOFTWARE: FastSeq for Windows Version 4.0					
: SEQ ID NO 1:					
: LENGTH: 2032					
: TYPE: DNA					
: ORGANISM: Homo sapiens					
US-09-045-284A-1					
Query Match					
Best Local Similarity 99.9% Pred. No. 0:					
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0:					
QY	106	CAAGCTCTTCACCTTCAGACACATGCTAGCTCTTAAAAAATGAAGCTCTGCTTTCT	165	Sequence 69, Appl	
DB	160	CAAGCTCTTCACCTTCAGACACATGCTAGCTCTTAAAAAATGAAGCTCTGCTTTCT	219	Sequence 69, Appl	
QY	166	GCTTCCAGATGGCCATCTGCTCTTATCTTCCATGATGACCAACATCAGCTC	225	Sequence 29, Appl	
DB	220	GCTTCCAGATGGCCATCTGCTCTTATCTTCCATGATGACCAACATCAGCTC	279	Sequence 29, Appl	
QY	226	CCTGTATGAGGACAGCCGAGCCGATGCTGCTGCTTCTGCTGCGCTC	285	Sequence 29, Appl	
DB	280	CCTGTATGAGGACAGCCGAGCCGATGCTGCTGCTTCTGCTGCGCTC	339	Sequence 29, Appl	
QY	286	TGCTCTTCTTCTTGGGGAGCTTTTGGGACACCCAGATCTTTTCTAGTATGA	345	Sequence 29, Appl	
DB	340	TGCTCTTCTTCTTGGGGAGCTTTTGGGACACCCAGATCTTTTCTAGTATGA	399	Sequence 29, Appl	
QY	346	GCCCGCTGAGCTGTGATGATCTTCAACAGACAGCCGCTGATGCTGACATGGC	405	Sequence 29, Appl	
DB	400	GCCCGCTGAGCTGTGATGATCTTCAACAGACAGCCGCTGATGCTGACATGGC	459	Sequence 29, Appl	
QY	406	TGTGCGGATCTGTATGAGGCGCTCTTCTTGTGACATGAGCTTTTGTATGCTACAT	465	Sequence 29, Appl	
DB	460	TGTGCGGATCTGTATGAGGCGCTCTTCTTGTGACATGAGCTTTTGTATGCTACAT	519	Sequence 29, Appl	
QY	466	GGAACCTGTCTCCCGGAGACAGTCCAGCTCTTTTCACTGAGGAGAAAGCGGCGCTGTG	525	Sequence 29, Appl	
DB					

```
Db 520 GGAACCTGGTCCCGGAGACATCGACGCTTTTCAGTGGGAGACAGCCGGGCTGTG 579
QY 526 TTCTGACCTGCGCTGTGACATCATCCACAAAGATGAATCATCCCCGGGCTCATCGAG 585
Db 580 TTCTGACCTGCGCTGTGACATCATCCACAAAGATGAATCATCCCCGGGCTCATCGAG 639
QY 586 GGTCCCTGTGAGTCAACACAGCCCTTTTGAAGTGGTGAAGAGGCTGCCGCTCTACAGCCA 645
Db 640 GCTCGTGTGAGTCAACACAGCCCTTTTGAAGTGGTGAAGAGGCTGCCGCTCTACAGCCA 699
QY 646 CGTGTGCTCAAGAGAGGTGGCTTTTCAACCTGCAGTCCCTTACCCGCTGCTGAAGA 705
Db 700 CGTGTGCTCAAGAGAGGTGGCTTTTCAACCTGCAGTCCCTTACCCGCTGCTGAAGA 759
QY 706 CCCCTCCCTCAACCTGCATATCGTGCACCTGGTCCGGACCCCGGGCGGTGTTCGCTTC 765
Db 760 CCCCTCCCTCAACCTGCATATCGTGCACCTGGTCCGGACCCCGGGCGGTGTTCGCTTC 819
QY 766 CCGAGAACCCACAAAGAGAGATCTCATGATTTGACATGCCATTTGTGATGGGCGACATGA 825
Db 820 CCGAGAACCCACAAAGAGAGATCTCATGATTTGACATGCCATTTGTGATGGGCGACATGA 879
QY 826 GCAAAAACCTCAAGAGAGAGACCAACCTACTATGTGATGACAGTGCATGCCAAAGCCA 885
Db 880 GCAAAAACCTCAAGAGAGAGACCAACCTACTATGTGATGACAGTGCATGCCAAAGCCA 939
QY 886 GCTGAGATCTACAGAACCATCCAGTCTTGGCCAAAGGCCCTGCAGAAAGCTTACCTGCT 945
Db 940 GCTGAGATCTACAGAACCATCCAGTCTTGGCCAAAGGCCCTGCAGAAAGCTTACCTGCT 999
QY 946 TTGCGCTATGAGAGACTGGCTGCAGCCCTGTGGCCAGACTTCCCGAATGTATGAAT 1005
Db 1000 TTGCGCTATGAGAGACTGGCTGCAGCCCTGTGGCCAGACTTCCCGAATGTATGAAT 1059
QY 1006 CGTGGATTTGGAATTTTCCCATCTTACAGACTGGTGGATGACATACATCAACCCGAGCAA 1065
Db 1060 CGTGGATTTGGAATTTTCCCATCTTACAGACTGGTGGATGACATACATCAACCCGAGCAA 1119
QY 1066 GGGCATGGTGAACACAGCTTTCACACAATGCGAGGATGCCCTTAATGTCTCCAGGC 1125
Db 1120 GGGCATGGTGAACACAGCTTTCACACAATGCGAGGATGCCCTTAATGTCTCCAGGC 1179
QY 1126 TTGGGCTGGTCTTGGCCCTATGAAAAGTTTCTGCACCTTCAGAAAAGCCTGTGGCATGC 1185
Db 1180 TTGGGCTGGTCTTGGCCCTATGAAAAGTTTCTGCACCTTCAGAAAAGCCTGTGGCATGC 1239
QY 1186 CATGAATTTGCTGGGCTACCGCACAGCTGATCTGAAGAAGACAGAGAAGCACTGTGCT 1245
Db 1240 CATGAATTTGCTGGGCTACCGCACAGCTGATCTGAAGAAGACAGAGAAGCACTGTGCT 1299
QY 1246 GGAATCTTGTCTACCTGAGCTGTGCTGAGCAAAATCACTAAGAGGTTGAGAAAGCTT 1305
Db 1300 GGAATCTTGTCTACCTGAGCTGTGCTGAGCAAAATCACTAAGAGGTTGAGAAAGCTT 1359
QY 1306 TGTGTGCACCTGGTGTGACGCTCAAGTCACTTTCTCGAATGCTTCTGAGACCTTGCTTAA 1365
Db 1360 TGTGTGCACCTGGTGTGACGCTCAAGTCACTTTCTCGAATGCTTCTGAGACCTTGCTTAA 1419
QY 1366 TGTGTGAGCCTTAACATACATGCTGTGGGTATCACACTGAGTGTGATGTGTCCACAG 1425
Db 1420 TGTGTGAGCCTTAACATACATGCTGTGGGTATCACACTGAGTGTGATGTGTCCACAG 1479
QY 1426 TGTCTAAGCAGAAAGACTTTTGTGTCCAGTGTGTGTCTAAGAAACAGAGTGGGAACT 1485
Db 1480 TGTCTAAGCAGAAAGACTTTTGTGTCCAGTGTGTGTCTAAGAAACAGAGTGGGAACT 1539
QY 1486 TATGTGAGCAGACATCCACAGTGAAGAGGATGCTCTCTCTCTCTCTCTCTCTCTCT 1545
Db 1540 TATGTGAGCAGACATCCACAGTGAAGAGGATGCTCTCTCTCTCTCTCTCTCTCTCTCT 1599
QY 1546 TCTGTGTGAGCAGATTCAGAGACTTTGTGGCTGAGAGGCTATTAAGCAGACAGACT 1605
Db 1600 TCTGTGTGAGCAGATTCAGAGACTTTGTGGCTGAGAGGCTATTAAGCAGACAGACT 1659
```

```
QY 1606 ATCAGTGAATTTGATCCATAAACCCTCCTGTCACATCTTGCCCAATGGGGAATGATCT 1665
Db 1660 ATCAGTGAATTTGATCCATAAACCCTCCTGTCACATCTTGCCCAATGGGGAATGATCT 1719
QY 1666 TTACCAAAAGAGCTCACACAGACTTTTCCACAGAGATGC 1703
Db 1720 TTACCAAAAGAGCTCACACAGACTTTTCCACAGAGATGC 1757

RESULT 2
US-09-190-911-2
; Sequence 2, Application US/09190911
; Patent No. 6365365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/09/190,911
; EARLIER FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens
US-09-190-911-2

Query Match 75.7%; Score 1547; DB 4; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 106 CAAGTCTTCCACTTGACAGCAATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTCT 165
Db 160 CAAGTCTTCCACTTGACAGCAATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTCT 219
QY 166 GCTTCCCAAGTGGCCATCTGGCTCTATTTCCACATGTACAGCCAAACATCAGCTC 225
Db 220 GCTTCCCAAGTGGCCATCTGGCTCTATTTCCACATGTACAGCCAAACATCAGCTC 279
QY 226 CCTGTATGAAGGACACAGCCGAGGAGCATGACGTGTGTCTTCTTCTGCGGCTC 285
Db 280 CCTGTATGAAGGACACAGCCGAGGAGCATGACGTGTGTCTTCTTCTGCGGCTC 339
QY 286 TGGCTCTCTTTTGTGGGCAAGCTTTTGGGACAGACCCAGATGTTTCTTACCTGATGA 345
Db 340 TGGCTCTCTTTTGTGGGCAAGCTTTTGGGACAGACCCAGATGTTTCTTACCTGATGA 399
QY 346 GCCCGCTGGACAGTGTGATGACCTTCAAGACAGACAGCCGCTGATGCTGACATGGC 405
Db 400 GCCCGCTGGACAGTGTGATGACCTTCAAGACAGACAGCCGCTGATGCTGACATGGC 459
QY 406 TGTGCGGATGTGATACGGGCGGCTTCTTGTGTGCAATGAGGCTTTGATGCTTACAT 465
Db 460 TGTGCGGATGTGATACGGGCGGCTTCTTGTGTGCAATGAGGCTTTGATGCTTACAT 519
QY 466 GGAACCTGGTCCCGGAGACAGTCCAGCTCTTTTCAAGTGGGAGAAACAGCCGGGCTGTG 525
Db 520 GGAACCTGGTCCCGGAGACAGTCCAGCTCTTTTCAAGTGGGAGAAACAGCCGGGCTGTG 579
QY 526 TTTGTGACCTGGCTGTGACATCATCCACAAAGATGAATCATCCCCGGGCTCATCGAG 585
Db 580 TTTGTGACCTGGCTGTGACATCATCCACAAAGATGAATCATCCCCGGGCTCATCGAG 639
QY 586 GCTCGTGTGACATCAACAGCCCTTTGAGGTGTGTGAAGAGGCTGCGCTCTTACAGCCA 645
Db 640 GCTCGTGTGACATCAACAGCCCTTTGAGGTGTGTGAAGAGGCTGCGCTCTTACAGCCA 699
```


QY 646 CGTGTGCTCAGAGAGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGCTGAAGA 705
DB 700 CGTGTGCTCAGAGAGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGCTGAAGA 759
QY 706 CCCCTCCCTCAACCTGATATCGTGCACCTGGTCCGGACCCCGGGCCGTTCCTTC 765
DB 760 CCCCTCCCTCAACCTGATATCGTGCACCTGGTCCGGACCCCGGGCCGTTCCTTC 819
QY 766 CCGAGAACGACAAAGGAGATGTCATGATGACAGTCCGATGATGATGGGAGCAGATGA 825
DB 820 CCGAGAACGACAAAGGAGATGTCATGATGACAGTCCGATGATGATGGGAGCAGATGA 879
QY 826 GCAAAAACCTAAGAGAGAGACCAACCCCTATGATGATGATGATGATGATGATGATGAT 885
DB 880 GCAAAAACCTAAGAGAGAGACCAACCCCTATGATGATGATGATGATGATGATGATGAT 939
QY 886 GCTGAGATCTACAGACACATCCAGTCTTGGCCCAAGGCTCGAGAGACGCTACCTGCT 945
DB 940 GCTGAGATCTACAGACACATCCAGTCTTGGCCCAAGGCTCGAGAGACGCTACCTGCT 999
QY 946 TGTGCGCTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTCCCGAATGATGAT 1005
DB 1000 TGTGCGCTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTCCCGAATGATGAT 1059
QY 1006 CGTGGGATTTGGAATTTTGGCCCATCTTCAAGACCTGGGTGATTAATCAACCCGAGCAA 1065
DB 1060 CGTGGGATTTGGAATTTTGGCCCATCTTCAAGACCTGGGTGATTAATCAACCCGAGCAA 1119
QY 1066 GGGCATGGTGACACGCTTCCACACAAATGCCAGGAGTCCCTTAATGCTCCAGGC 1125
DB 1120 GGGCATGGTGACACGCTTCCACACAAATGCCAGGAGTCCCTTAATGCTCCAGGC 1179
QY 1126 TTTGGCGCTGCTTGTGGCTATGAAAAGTTTCTGACTTCAGAAACCTTGGCGATGC 1185
DB 1180 TTTGGCGCTGCTTGTGGCTATGAAAAGTTTCTGACTTCAGAAACCTTGGCGATGC 1239
QY 1186 CATGAATTTCTGGGCTACCGCCACGTGATGTAACAGAACAGAGAAACCTGTTGCT 1245
DB 1240 CATGAATTTCTGGGCTACCGCCACGTGATGTAACAGAACAGAGAAACCTGTTGCT 1299
QY 1246 GGAATCTTCTGTACCTGAGATGTCCTGAGCAATTCACATAAGAGGTTGAGAAAGCTT 1305
DB 1300 GGAATCTTCTGTACCTGAGATGTCCTGAGCAATTCACATAAGAGGTTGAGAAAGCTT 1359
QY 1306 TGTGCGACCTGCTGAGCTCAGTCACTTCTCTGAATGCTTCTGAGCTTCCCTACA 1365
DB 1360 TGTGCGACCTGCTGAGCTCAGTCACTTCTCTGAATGCTTCTGAGCTTCCCTACA 1419
QY 1366 TCTCTGAGCTTAACTACATGCTGTGGGTATCACTGATGATGATGATGATGATGATGAT 1425
DB 1420 TCTCTGAGCTTAACTACATGCTGTGGGTATCACTGATGATGATGATGATGATGATGAT 1479
QY 1426 TGTCTAAGCAAGAGAGATTTTGTGTCCATGCTGTGTGTAGAAAACAGACTGGGAACT 1485
DB 1480 TGTCTAAGCAAGAGAGATTTTGTGTCCATGCTGTGTGTGTAGAAAACAGACTGGGAACT 1539
QY 1486 TATGTGAGCAGACATCCCAAGTGAAGAGGATTTGCTCTTCTTCTTGTGATCT 1545
DB 1540 TATGTGAGCAGACATCCCAAGTGAAGAGGATTTGCTCTTCTTCTTGTGATCT 1599
QY 1546 TCTCTGTGGGAGACTTTCAGAGACTTGTGGCTCGAGAGGCTTATTAAGCAGCAGAGT 1605
DB 1600 TCTCTGTGGGAGACTTTCAGAGACTTGTGGCTCGAGAGGCTTATTAAGCAGCAGAGT 1659
QY 1606 ATCAGTGAATTTGATCAATAAAGCTCCCTGTCACATCTTCCCAATGGGGAATGATCT 1665
DB 1660 ATCAGTGAATTTGATCAATAAAGCTCCCTGTCACATCTTCCCAATGGGGAATGATCT 1719
QY 1666 TTCACCAAGAGCTCAGCAGATTTTCACAGAGATGCT 1703
DB 1720 TTCACCAAGAGCTCAGCAGATTTTCACAGAGATGCT 1757

RESULT 3
US-09-786-240-26
Sequence 26 Application US/09786240
Patent No. 6558935
GENERAL INFORMATION:
APPLICANT: INCYTE PHARMACEUTICALS, INC.
APPLICANT: TANG, Y. Tom
APPLICANT: CORLEY, Neil C.
APPLICANT: GUEGLER, Karl J.
APPLICANT: BAUGHN, Mariah R.
APPLICANT: LAL, Preeti
APPLICANT: YDE, Henry
APPLICANT: HILMAN, Jennifer L.
APPLICANT: AZIMZAI, Yalda
TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
FILE REFERENCE: PF-0592 PCT
CURRENT APPLICATION NUMBER: US/09/786,240
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 09/150,657; unaassigned; 09/186,779; unaassigned; 60/133,6
PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1998-11-04; 1999-05-11
SOFTWARE: PERL Program
SEQ ID NO 26
LENGTH: 2065
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6558935 2617407CB1
US-09-786-240-26

Query Match 56.4%; Score 1153; DB 4; Length 2065;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1593; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 106 CAAGTCTTCCACTTCAGACACATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTCT 165
DB 152 CAAGTCTTCCACTTCAGACACATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTCT 211
QY 166 GGTTCACAGATGAGCAATCTTGGCTTATTTCTTCCATGATGACGACCAACATCAGCTC 225
DB 212 GGTTCACAGATGAGCAATCTTGGCTTATTTCTTCCATGATGACGACCAACATCAGCTC 271
QY 226 CCTGTATGAAAGCAGACGCCAGCGCATCAGTGTGCTGCTTCTTCTGCGCTC 285
DB 272 CCTGTATGAAAGCAGACGCCAGCGCATCAGTGTGCTGCTTCTTCTGCGCTC 331
QY 286 TGGCTCTTCTTGTGGGGAGCTTTTGGGACACCAAGATGTTTCTACGTGATGA 345
DB 332 TGGCTCTTCTTGTGGGGAGCTTTTGGGACACCAAGATGTTTCTACGTGATGA 391
QY 346 GCCCGCTGGACGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 405
DB 392 GCCCGCTGGACGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 451
QY 406 TGTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 465
DB 452 TGTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 511
QY 466 GGAACCTGTGCTCCCGGAGACAGTCCAGGCTTTTCACTGAGGAGAACACCGGGCCCTGTG 525
DB 512 GGAACCTGTGCTCCCGGAGACAGTCCAGGCTTTTCACTGAGGAGAACACCGGGCCCTGTG 571
QY 526 TTTCTGACCTTCTGTCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 585
DB 572 TTTCTGACCTTCTGTCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 630
QY 586 GCTCTGTGAGTAAAGAGCTTTG--AGGTGGTGAAGAGCTGCGCTCTACAGCC 644
DB 631 GCTCTGTGAGTAAAGAGCTTTGAGATTTTGGAGAAAGCTTGGCTCTACAGCC 690
QY 645 ACGTGTGCTCAGAGAGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGCTGAAG 704

STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/03034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3792 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: YES
IMMEDIATE SOURCE:
CLONE: pg-host4
US-08-992-334-1

Query Match 0.9%; Score 19; DB 2; Length 3792;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCTGTGACAAAGAGCAAG 1839
Db 1818 TCTGTGACAAAGAGCAAG 1836

RESULT 7
US-08-302-752-1
Sequence 1, Application US/08302752
Patent No. 6025190
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: THERMOSENSIBLE PLASMID
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3792 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-1

Query Match 0.9%; Score 19; DB 3; Length 3792;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCTGTGACAAAGAGCAAG 1839
Db 1818 TCTGTGACAAAGAGCAAG 1836

RESULT 8
US-08-992-334-2
Sequence 2, Application US/08992334
Patent No. 5919678
GENERAL INFORMATION:
APPLICANT: Gruss, Alexandra
APPLICANT: Maguin, Emmanuelle
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christie Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/03034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5234 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-992-334-2

Query Match 0.9%; Score 19; DB 2; Length 5234;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1821 TCCTGCAAAAGAGCAAG 1839
DB 3260 TCCTGCAAAAGAGCAAG 3278

RESULT 9
US-08-302-752-2

; Sequence 2, Application US/08302752
; Patent No. 6025190
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: THERMOSENSIBLE PLASMID
; NUMBER OF SEQUENCES: 3
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (PPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,752
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9203034
; FILING DATE: 13-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO FR/93/00248
; FILING DATE: 12-MAR-1993
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-302-752-2

Query Match 0.9%; Score 19; DB 3; Length 5234;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1821 TCCTGCAAAAGAGCAAG 1839
DB 3260 TCCTGCAAAAGAGCAAG 3278

RESULT 10
US-08-992-334-3

; Sequence 3, Application US/08992334
; Patent No. 5919678
; GENERAL INFORMATION:
; APPLICANT: Gruss, Alexandra
; APPLICANT: Masulin, Emmanuelle
; TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christie Parker & Hale, LLP
; STREET: 350 West Colorado Boulevard, Suite 500
; CITY: Pasadena
; STATE: California
; COUNTRY: United States
; ZIP: 91105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/992,334
; FILING DATE: 17-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/302,752
; FILING DATE: 24-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00248
; FILING DATE: 12-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/03034
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Prout, D. Bruce
; REGISTRATION NUMBER: 20958
; REFERENCE/DOCKET NUMBER: C93:31779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 795-9900
; TELEFAX: (626) 577-8800
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6722 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
US-08-992-334-3

Query Match 0.9%; Score 19; DB 2; Length 6722;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1821 TCCTGCAAAAGAGCAAG 1839
DB 4748 TCCTGCAAAAGAGCAAG 4766

RESULT 11
US-08-302-752-3

; Sequence 3, Application US/08302752
; Patent No. 6025190
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: THERMOSENSIBLE PLASMID
; NUMBER OF SEQUENCES: 3
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (PPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,752
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9203034
; FILING DATE: 13-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO FR/93/00248
; FILING DATE: 12-MAR-1993
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6722 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-302-752-3

Query Match 0.9%; Score 19; DB 3; Length 6722;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1821 TCCTGCAAAAGAGCAAG 1839
DB 4748 TCCTGCAAAAGAGCAAG 4766

```
RESULT 12
US-08-778-494B-112
; Sequence 112, Application US/08778494B
; Patent No. 5962272
; GENERAL INFORMATION:
; APPLICANT: Chenchik, Alex
; APPLICANT: Zhu, York
; APPLICANT: Diachenko, Luda
; APPLICANT: Siebert, Paul
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR FULL-LENGTH CDNA
; TITLE OF INVENTION: CLONING
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/778,494B
; FILING DATE: 03-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/582,562
; FILING DATE: 03-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Doran R.
; REGISTRATION NUMBER: 38,261
; REFERENCE/DOCKET NUMBER: CL-7C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (352) 375-8100
; TELEFAX: (352) 372-5800
; INFORMATION FOR SEQ ID NO: 112:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (synthetic)
US-08-778-494B-112

Query Match          0.9%; Score 18; DB 2; Length 23;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      21 CGGAAGCAGCAGAGGG 38
        |||
        6 CGGAAGCAGCAGAGGG 23

RESULT 13
US-09-045-284A-8/C
; Sequence 8, Application US/09045284A
; Patent No. 6265192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerlich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107U51
; CURRENT APPLICATION NUMBER: US/09/045,284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 47
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-045-284A-8

Query Match          0.9%; Score 18; DB 3; Length 47;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1248 ATCTTCGTCTACCTGGA 1265
        |||
        47 ATCTTCGTCTACCTGGA 30

RESULT 14
US-09-190-911-8/C
; Sequence 8, Application US/09190911
; Patent No. 6365365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerlich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/09/190,911
; EARLIER FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 47
; TYPE: DNA
; ORGANISM: H. sapiens
US-09-190-911-8

Query Match          0.9%; Score 18; DB 4; Length 47;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1248 ATCTTCGTCTACCTGGA 1265
        |||
        47 ATCTTCGTCTACCTGGA 30

RESULT 15
US-09-439-313-320
; Sequence 320, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yugu
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Rietter, Mark
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 320
; LENGTH: 150
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-439-313-320
```

Query Match 0.98; Score 18; DB 4; Length 150;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 GAATCCATTGTTGGG 18
|||||
DB 31 GAATCCATTGTTGGG 48

Search completed: August 15, 2003, 00:46:53
Job time : 143 secs

6

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 20:11:42 ; Search time 7280 Seconds

(without alignments)
11480.545 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043

Sequence: 1 gaattccattgtgtgtgta.....tgggattccctaaaaaaa 2043

Scoring table: OLIGO_NMC

Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 2045481386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5682662

Post-processing: Listing first 45 summaries

Database : GenEmbl:

1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_ov:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_rtd:*
37: em_htg_rtd:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1722	84.3	1992	9	AF280088	AF280088 Homo sapi
2	1653	80.9	2037	9	BC035282	BC035282 Homo sapi
3	1598	78.2	2011	9	AK026635	AK026635 Homo sapi
4	1597	78.2	183228	9	AC010547	AC010547 Homo sapi
5	1597	78.2	206943	2	AC138848	AC138848 Homo sapi
6	1547	75.7	1979	6	BD127258	BD127258 Homo sapi
7	1547	75.7	1979	6	AK074746	AK074746 Homo sapi
8	1547	75.7	2032	6	AR203335	AR203335 Sequence
9	1547	75.7	2032	6	BD134772	BD134772 Glycosyls
10	1547	75.7	2032	6	AF131235	AF131235 Homo sapi
11	1262	61.8	1333	9	AF149783	AF149783 Homo sapi
12	1153	56.4	2065	6	AR310485	AR310485 Sequence
13	695	34.0	877	6	BD124832	BD124832 Primer fo
14	695	34.0	877	6	BD126950	BD126950 Primer fo
15	403	19.7	517	6	AX381256	AX381256 Sequence
16	114	5.6	389	6	AX381326	AX381326 Sequence
17	105	5.1	548	6	BD125617	BD125617 Primer fo
18	59	2.9	1462	9	AF176839	AF176839 Homo sapi
19	59	2.9	1647	6	AX327330	AX327330 Sequence
20	59	2.9	2170	9	AF176838	AF176838 Homo sapi
21	59	2.9	2544	9	AF219990	AF219990 Homo sapi
22	59	2.9	3278	9	AF246718	AF246718 Homo sapi
23	59	2.9	3786	9	AF280086	AF280086 Homo sapi
24	59	2.9	71503	9	AF219991	AF219991 Homo sapi
25	59	2.9	157358	9	AC009163	AC009163 Homo sapi
26	59	2.9	194832	9	AC025287	AC025287 Homo sapi
27	59	2.9	208185	2	AC009105	AC009105 Homo sapi
28	37	1.8	37	6	AR203340	AR203340 Sequence
29	37	1.8	37	6	BD134774	BD134774 Glycosyls
30	30	1.5	30	6	BD134553	BD134553 Method fo
31	31	1.3	1926	6	BD134773	BD134773 Glycosyls
32	27	1.3	1926	10	AF131236	AF131236 Mus muscu
33	27	1.3	2201	10	AF109155	AF109155 Mus muscu
34	27	1.3	232745	2	AC130174	AC130174 Rattus no
35	27	1.3	262679	2	AC110314	AC110314 Rattus no
36	24	1.2	228569	2	AC121740	AC121740 Rattus no
37	24	1.2	235173	2	AC133449	AC133449 Rattus no
38	24	1.2	240901	2	AC096328	AC096328 Rattus no
39	24	1.2	242253	2	AC095651	AC095651 Rattus no
40	24	1.2	246877	2	AC137365	AC137365 Rattus no
41	24	1.2	258097	2	AC129764	AC129764 Rattus no
42	24	1.2	271230	2	AC094919	AC094919 Rattus no
43	23	1.1	1740	10	AF176841	AF176841 Mus muscu
44	23	1.1	1989	10	AF176840	AF176840 Mus muscu
45	23	1.1	110000	2	AC099262_1	Continuation (2 of

ALIGNMENTS

RESULT 1
AF280088
LOCUS
DEFINITION Homo sapiens L-selectin ligand sulfotransferase GST-3 mRNA,
complete cds.
ACCESSION AF280088
VERSION AF280088.1 GI:12060807
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1992)
Hemmerich,S., Lee,J.K., Bhakta,S., Bistrup,A., Ruddle,N.R. and
Rosen,S.D.

TITLE Chromosomal localization and genomic organization for the
 galactose/N-acetylglucosamine/N-acetylglucosamine
 6-O-sulfotransferase gene family
JOURNAL Glycobiology 11 (1), 75-87 (2001)
MEDLINE 21096027
PUBMED 11181564
REFERENCE 2 (bases 1 to 1992)
AUTHORS Hemmerlich, S., Bhakta, S., Lee, J.-K., Bistrup, A., Ruddle, N.R. and
 Rosen, S.D.
TITLE Direct Submision
JOURNAL Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience,
 3401 Hillview Ave., Palo Alto, CA 94304, USA
FEATURES Location/Qualifiers
 source
 1..1992
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="16"
 /map="16q23.1-q23.2"
 /cell_type="high endothelial"
 /tissue_type="Tonsil"
 1..1992
 /gene="GST3"
 122..1282
 /note="similar to Homo sapiens sequence encoded by GenBank
 Accession Number AF131235"
 /product="L-selectin ligand sulfotransferase GST-3"
 /protein_id="AA048246.1"
 /db_xref="GI:12060808"
 /translation="MILPKKMLLFLVSQMALFLFPHSHNLSLSMKAPKPMH
 VVLSMRSGSFFVGOLFQHPDVFILMEPAHWVTFKOSTAMILHMAVRLIRAVE
 LCDMSVFDAYMEGPRROSLPQWENSRLCSAPCDIIPDEIIPRAKRLISQOP
 FEYVAKRSYSHVYLKEVFENLQSLYLDKPSILHLIVLRPAVFRSREKTP
 GMLIDSRIVMGHOEKLKEDOPYVMOVICOSOLEIKTLOSPLKAOEYELVRY
 EDLAPAVQTSRMTEFVGLLEFLPHLOTFVHNITRGKMGDIAFPHNADALVNSQAM
 RNSLEYKYSRLOKAGDMANLIGTRHRSBEDQRMLDLSTWTPQIHT"
 BASE COUNT 482 a 540 c 489 g 481 t
 ORIGIN
 Query Match 84.3%; Score 1722; DB 9; Length 1992;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1722; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 427 CGCTCTTGTGCGACATGAGCGCTCTTGATGCTCATGATGAACTGTGCCCCGAGACA 486
 DB 421 CGCTCTTGTGCGACATGAGCGCTCTTGATGCTCATGATGAACTGTGCCCCGAGACA 480
 QY 487 GTCCAGCCCTTTCATGATGGAGAACAGCCGGGCGCTGTGTGCACTGCTGTGACAT 546
 DB 481 GTCCAGCCCTTTCATGATGGAGAACAGCCGGGCGCTGTGTGCACTGCTGTGACAT 540
 QY 547 CATCCCAAGATGAAATCATCCCGGGGCTCACTGACAGCTCTGTGACATCAACAGCC 606
 DB 541 CATCCCAAGATGAAATCATCCCGGGGCTCACTGACAGCTCTGTGACATCAACAGCC 600
 QY 607 CTTTGAGTGTGAGAACAGCCCTGCCCTCTCAACGCCACGTGTGTCTCAAGAGGTGCG 666
 DB 601 CTTTGAGTGTGAGAACAGCCCTGCCCTCTCAACGCCACGTGTGTCTCAAGAGGTGCG 660
 QY 667 CTTCTCAACCTGACAGTCCCTGTACCCCGCTGTGTAAGAACCCCTCCCTCAACCTGATAT 726
 DB 661 CTTCTCAACCTGACAGTCCCTGTACCCCGCTGTGTAAGAACCCCTCCCTCAACCTGATAT 720
 QY 727 CGTGCACTGTGTCGGGACCCCGGGCGCTGTTCGTTCCGAGAACGCAAAAGGAGAGA 786
 DB 721 CGTGCACTGTGTCGGGACCCCGGGCGCTGTTCGTTCCGAGAACGCAAAAGGAGAGA 780
 QY 787 TCTCATGATTTGACAGTCGATTTGTATGGGGGACATGACAAAAAATCAAGAGAGAGA 846
 DB 781 TCTCATGATTTGACAGTCGATTTGTATGGGGGACATGACAAAAAATCAAGAGAGAGA 840
 QY 847 CCACACCTCTATGATGATCAGATCATGACCAAAAGCCAGCGAGATCTACAAAGCAT 906
 DB 841 CCACACCTCTATGATGATCAGATCATGACCAAAAGCCAGCGAGATCTACAAAGCAT 900
 QY 907 CCAGTCTTGCCCAAGGCGCTCCAGAGAACGCTACCTGCTTGCGCTATGAGACCTGCG 966
 DB 901 CCAGTCTTGCCCAAGGCGCTCCAGAGAACGCTACCTGCTTGCGCTATGAGACCTGCG 960
 QY 967 TCGAGCCCTGTGCGCCCAACTCCCGATGTATGTAATTCGTGGGATTTGGAATTTCTGCG 1026
 DB 961 TCGAGCCCTGTGCGCCCAACTCCCGATGTATGTAATTCGTGGGATTTGGAATTTCTGCG 1020
 QY 1027 CCATCTTACAGCTGGGTGATATCATACCTCCGAGGCAAGGCGATGAGGACACGCTTT 1086
 DB 1021 CCATCTTACAGCTGGGTGATATCATACCTCCGAGGCAAGGCGATGAGGACACGCTTT 1080
 QY 1087 CCACACAATGCGCAGGATGCCCTTAATGTCTCCAGGCTTGCGCTGCTTTGCCCTTA 1146
 DB 1081 CCACACAATGCGCAGGATGCCCTTAATGTCTCCAGGCTTGCGCTGCTTTGCCCTTA 1140
 QY 1147 TGAAGAGTTTCTCCAGCTTCAAGAAAGCCGTGTGGCATGCGATGAAATTTGCTGGCTACCG 1206
 DB 1141 TGAAGAGTTTCTCCAGCTTCAAGAAAGCCGTGTGGCATGCGATGAAATTTGCTGGCTACCG 1200
 QY 1207 CCACGTCAGATCTGAACAAGAACAGAAACCTGTTGCGATCTTCTGTCTACCTGTGAC 1266
 DB 1201 CCACGTCAGATCTGAACAAGAACAGAAACCTGTTGCGATCTTCTGTCTACCTGTGAC 1260
 QY 1267 TGTCCCTGAGCAATCCACTAAGAGGTTGAGAAGGCTTTGTGCGACCTGTGTGACGC 1326
 DB 1261 TGTCCCTGAGCAATCCACTAAGAGGTTGAGAAGGCTTTGTGCGACCTGTGTGACGC 1320
 QY 1327 TTAGTCACCTTCTCTGAATGCTTCTGAGACCTTGCTTAATCTCTAGAGCTTAACATCATG 1386
 DB 1321 TTAGTCACCTTCTCTGAATGCTTCTGAGACCTTGCTTAATCTCTAGAGCTTAACATCATG 1380
 QY 1387 TGTGTGGGATACACTGAGTGTGATGTGTGTCACAGGTGTCGAAGAGAAAGACTTTT 1446
 DB 1381 TGTGTGGGATACACTGAGTGTGATGTGTGTCACAGGTGTCGAAGAGAAAGACTTTT 1440
 QY 1447 GTGTCCATGCTTGTGTCTAGAAACAGACTGGGGAACCTTATGTGAGAGCAATCCAC 1506
 DB 1441 GTGTCCATGCTTGTGTCTAGAAACAGACTGGGGAACCTTATGTGAGAGCAATCCAC 1500
 QY 1507 CAGTGAACAGGATATGCTCTTCTTCTTCTGTGATCTTCTCTGCTGTGGGACACTTACG 1566


```

OY 951 GGTATGAGACCTGGCTGAGCCCTGTGTGCCCCAGACTTCCGATGATGTAATTCGTGG 1010
Db 901 GGTATGAGACCTGGCTGAGCCCTGTGTGCCCCAGACTTCCGATGATGTAATTCGTGG 960
OY 1011 GATTGAATTTTGGCCCATTTTCAGACCTGGGTGATTAACATCCAGGAGGAGGGA 1070
Db 961 GATTGAATTTTGGCCCATTTTCAGACCTGGGTGATTAACATCCAGGAGGAGGGA 1020
OY 1071 TGGGTGACACGCTTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTTGGC 1130
Db 1021 TGGGTGACACGCTTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTTGGC 1080
OY 1131 GGTGGCTTTTGGCCCATTAAGAAAGCTTTCAGACTTCAGAAAGCTTGGCATCATTA 1190
Db 1081 GGTGGCTTTTGGCCCATTAAGAAAGCTTTCAGACTTCAGAAAGCTTGGCATCATTA 1140
OY 1191 ATTGCTGGGCTACCGCCACGTCAGATCTGAACAAGAAAGAAACCTGTGCTGGATC 1250
Db 1141 ATTGCTGGGCTACCGCCACGTCAGATCTGAACAAGAAAGAAACCTGTGCTGGATC 1200
OY 1251 TTCTGTACACCTGACCTGCTCCCTGAGCAAAATCCACTAAGAGGTTGAGAAGCTTGTCTG 1310
Db 1201 TTCTGTACACCTGACCTGCTCCCTGAGCAAAATCCACTAAGAGGTTGAGAAGCTTGTCTG 1260
OY 1311 CCACCTGGTGTACGCTCAGTCACTTCTCTGAATGCTTCTGAAGCTTCCCTACATCTCT 1370
Db 1261 CCACCTGGTGTACGCTCAGTCACTTCTCTGAATGCTTCTGAAGCTTCCCTACATCTCT 1320
OY 1371 GAGCCTTAACATAGTGTGTGGTATCAGACTGAGTGTGATGTGTGTCACACGCTC 1430
Db 1321 GAGCCTTAACATAGTGTGTGGTATCAGACTGAGTGTGATGTGTGTCACACGCTC 1380
OY 1431 AAGCAGAAAGACTTTTGTGTCCATGCTGTGTCTAGAAAACAGACTGGGAACTTATCT 1490
Db 1381 AAGCAGAAAGACTTTTGTGTCCATGCTGTGTCTAGAAAACAGACTGGGAACTTATCT 1440
OY 1491 GAGCAGACATCCCAACAGATGAAGGATTTGCTCTTCTTCTTGTGATCTTCTG 1550
Db 1441 GAGCAGACATCCCAACAGATGAAGGATTTGCTCTTCTTCTTGTGATCTTCTG 1500
OY 1551 TCTGGGACAGACTTCAGAGACTTGTGGCTGAGGAGGCTTAAAGCAGACAGATATCAG 1610
Db 1501 TCTGGGACAGACTTCAGAGACTTGTGGCTGAGGAGGCTTAAAGCAGACAGATATCAG 1560
OY 1611 TEGAAATTGATCCATAAACCCTCCCTGTCCACATCTTGCCCAATGGGGAATGATCTTTCAC 1670
Db 1561 TEGAAATTGATCCATAAACCCTCCCTGTCCACATCTTGCCCAATGGGGAATGATCTTTCAC 1620
OY 1671 CAAAGAGCTCAGCAGATTTTCCACAGAGATGC 1703
Db 1621 CAAAGAGCTCAGCAGATTTTCCACAGAGATGC 1653

```

```

RESULT 3
ACCESSION AK026635 2011 bp mRNA linear PRI 29-SEP-2000
LOCUS Homo sapiens cDNA: FLJ22982 fis, clone KAT11454, highly similar to
DEFINITION AF131235 Homo sapiens N-acetylglucosamine 6-O-sulfotransferase
mRNA.

```

```

ACCESSION AK026635.1 GI:10439531
VERSION AK026635.1
KEYWORDS oligo cloning, fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

```

```

REFERENCE 1 (sites)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
Matsunabe, K., Kumagai, A., Itakura, S., Yamazaki, M., Tashiro, H.,
Ota, T., Suzuki, Y., Okeyashiki, M., Nishi, T., Shibahara, T., Tanaka, T.,
Nakamura, Y., Isogai, T. and Sugano, S.

```

```

TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2011)

```

```

AUTHORS Sugano, S., Suzuki, Y., Ota, T., Okeyashiki, M., Nishi, T., Isogai, T.,
Shibahara, T., Tanaka, T. and Nakamura, Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
Japan (E-mail: cdna@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286,
Fax: 81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'-3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source location/Qualifiers
1..2011
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="KAT11454"
/cell_line="KATO III"
/cell_type="signed-ring cell carcinoma"
/clone_1kb="KAT"
/notes="cloning vector pME18SFL3"
/notes="highly similar to AF131235 Homo sapiens
N-acetylglucosamine 6-O-sulfotransferase mRNA"
BASE COUNT 482 a 541 c 475 g 513 t
ORIGIN
Query Match 78.2%; Score 1598; DB 9; Length 2011;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 1598; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 106 CAAGCTCTCCACTTACAGCAATGCTACTGCTTAAAAAAGAACTCTGCTGTTCT 165
Db 50 CAAGCTCTCCACTTACAGCAATGCTACTGCTTAAAAAAGAACTCTGCTGTTCT 109
OY 166 GATTTCACAGATGGCCATTTGGCTATCTTCCATGATACAGCAGCAATCAGCTC 225
Db 110 GATTTCACAGATGGCCATTTGGCTATCTTCCATGATACAGCAGCAATCAGCTC 169
OY 226 CCTGTCTAAGAGGACAGCCGAGCGCATGCACTGCTGTCTCTCTGCGGCTC 285
Db 170 CCTGTCTAAGAGGACAGCCGAGCGCATGCACTGCTGTCTCTCTGCGGCTC 229
OY 286 TGGCTCTCTTTTGGGGGACGCTTTTGGGAGCAGCCAGATGTTTCTACCTGATGA 345
Db 230 TGGCTCTCTTTTGGGGGACGCTTTTGGGAGCAGCCAGATGTTTCTACCTGATGA 289
OY 346 GCCCGCTGGCAGCTGTGATGACCTTCAAGCAGAGACAGCCGCTGATGCTGACATGCG 405
Db 290 GCCCGCTGGCAGCTGTGATGACCTTCAAGCAGAGACAGCCGCTGATGCTGACATGCG 349
OY 406 TGTGGGAGATGATACGCGCCGCTCTTCTTGTGCGACATGAGCGCTTGTGATGCTACT 465
Db 350 TGTGGGAGATGATACGCGCCGCTCTTCTTGTGCGACATGAGCGCTTGTGATGCTACT 409
OY 466 GGAACCTGGTCCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAACAGCCGGGCGCTGG 525
Db 410 GGAACCTGGTCCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAACAGCCGGGCGCTGG 469
OY 526 TTCTGACCTGCTGTGATCATCCACAGATGAATCATCCCGGCGTCACTGACG 585
Db 470 TTCTGACCTGCTGTGATCATCCACAGATGAATCATCCCGGCGTCACTGACG 529
OY 586 GCTCTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCTCTACAGCA 645
Db 530 GCTCTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCTCTACAGCA 589
OY 646 GGTGGTCAAGAGGAGGCTGCTTCAACTGAGTCCCTACCGGCTGCTGAAGA 705

```

Db 590 CGTGTGCTCAGAGAGTGGCTTCTTCAACCTGACAGTCCCTCTACCCGCTGCTGANAAGA 649
 QY 706 CCCCTCCCTACACTGATATGCGACCTGCTCCGAGACCCCGGCGGCTGCTGCTTC 765
 Db 650 CCCCTCCCTACACTGATATGCTGACCTGCTCCGAGACCCCGGCGGCTGCTGCTTC 709
 QY 766 CCGAGAGACGACAAAGGAGATCTCATGATGTGACAGTCCGATTTGATGGGCGAGCATGA 825
 Db 710 CCGAGAGACGACAAAGGAGATCTCATGATGTGACAGTCCGATTTGATGGGCGAGCATGA 769
 QY 826 GCAAAAACCTAAGAGAGAGACCAACCCCTACTATGTATGATGAGTCTTCCCAAAGCCA 885
 Db 770 GCAAAAACCTAAGAGAGAGACCAACCCCTACTATGTATGATGAGTCTTCCCAAAGCCA 829
 QY 886 GCTGAGATCTACAAAGACATCCAGTCTTCCCAAGGCGCTGAGAGAGCGCTACTGCT 945
 Db 830 GCTGAGATCTACAAAGACATCCAGTCTTCCCAAGGCGCTGAGAGAGCGCTACTGCT 889
 QY 946 TGTGCGCTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTTCCGAATGTATGAAT 1005
 Db 890 TGTGCGCTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTTCCGAATGTATGAAT 949
 QY 1006 CGTGGGATTTGGAATTTCTTGGCCCATTTTCAGACCTGGGTGCTATACATCACCCGAGCAA 1065
 Db 950 CGTGGGATTTGGAATTTCTTGGCCCATTTTCAGACCTGGGTGCTATACATCACCCGAGCAA 1009
 QY 1066 GGGCATGGTGACACCGCTTTCACACAAATGCCAGGATGCCCTTATGTCTCCAGGC 1125
 Db 1010 GGGCATGGTGACACCGCTTTCACACAAATGCCAGGATGCCCTTATGTCTCCAGGC 1069
 QY 1126 TTGGCGCTGTCTTTGGCCCTATGAGAAAGTTTCTGACTTCAGAAAGCCTGTGGCGATGC 1185
 Db 1070 TTGGCGCTGTCTTTGGCCCTATGAGAAAGTTTCTGACTTCAGAAAGCCTGTGGCGATGC 1129
 QY 1186 CATGAATTTGCTGGGCTACCGCCAGCTCATGATCGAACAAGAGAAACCTGTGCT 1245
 Db 1130 CATGAATTTGCTGGGCTACCGCCAGCTCATGATCGAACAAGAGAAACCTGTGCT 1189
 QY 1246 GGAATCTTCTGTACTGACTGCTGCTGAGCAATCCACTAAGAGGTTGAGAGGCTT 1305
 Db 1190 GGAATCTTCTGTACTGACTGCTGCTGAGCAATCCACTAAGAGGTTGAGAGGCTT 1249
 QY 1306 TGTGCGCACTGGTGTACGCTCAGTCACTTCTGTAATGCTTCTGAGCCTTCCCTACA 1365
 Db 1250 TGTGCGCACTGGTGTACGCTCAGTCACTTCTGTAATGCTTCTGAGCCTTCCCTACA 1309
 QY 1366 TGTGAGCCTTACTACATGCTGCTGGGTATCATCACTGATGATGCTGCTCCACAG 1425
 Db 1310 TGTGAGCCTTACTACATGCTGCTGGGTATCATCACTGATGATGCTGCTCCACAG 1369
 QY 1426 TGTCAAGCAGAGAGACTTTTGTGTCATGCTTGTCTAGAAACAGACTGGGGAACCT 1485
 Db 1370 TGTCAAGCAGAGAGACTTTTGTGTCATGCTTGTCTAGAAACAGACTGGGGAACCT 1429
 QY 1486 TATGTGAGCAGACATCCACAGTGAAGAGGGTATGCTCTTCTTCTTCTGATCT 1545
 Db 1430 TATGTGAGCAGACATCCACAGTGAAGAGGGTATGCTCTTCTTCTTCTGATCT 1489
 QY 1546 TCTGTGTGGGAGACTTCAAGAGCTTGTGGCTGAGAGGCTTAAAGCAGCAGACAGT 1605
 Db 1490 TCTGTGTGGGAGACTTCAAGAGCTTGTGGCTGAGAGGCTTAAAGCAGCAGACAGT 1549
 QY 1606 ATCAGTGAATTTGATATTAACCTCCCTGTGCCACATCTTGGCCAAATGGGGAATGATCT 1665
 Db 1550 ATCAGTGAATTTGATATTAACCTCCCTGTGCCACATCTTGGCCAAATGGGGAATGATCT 1609
 QY 1666 TTCACAAAGAGCTCACAGACTTTTCCACAGAGATGC 1703
 Db 1610 TTCACAAAGAGCTCACAGACTTTTCCACAGAGATGC 1647

LOCUS AC010547 183228 bp DNA linear PRI 29-SEP-2001
 DEFINITION Homo sapiens chromosome 16 clone RP11-510M2, complete sequence.
 AC010547
 AC010547.9 GI:15808510
 VERSION
 AC010547.9
 KEYWORDS
 HTG.
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 183228)
 DOE Joint Genome Institute.
 Sequencing of Human Chromosome 16
 Unpublished
 REFERENCE
 2 (bases 1 to 183228)
 DOE Joint Genome Institute.
 Direct Submission
 TITLE
 Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE
 3 (bases 1 to 183228)
 DOE Joint Genome Institute.
 Direct Submission
 TITLE
 Submitted (29-SEP-2001) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 On Sep 29, 2001 this sequence version replaced gi:14589436.
 COMMENT
 Sequence Quality Assessment:
 This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 Genbank flat file format but are available as part
 of this entry's ASN.1 file.

 Sequence Quality Assessment:
 This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 Genbank flat file format but are available as part
 of this entry's ASN.1 file.

 FEATURES
 source
 1. 183228
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-510M2"
 BASE COUNT 51901 a 42098 c 40026 g 49203 t
 ORIGIN
 Query Match 78.2%; Score 1597; DB 9; Length 183228;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 107 AAGCTTCCACTTCAGACATGCTACTGCTTAAAAAATGAAGCTCTCTGTTCTG 166
 Db 142584 AAGCTTCCACTTCAGACATGCTACTGCTTAAAAAATGAAGCTCTCTGTTCTG 142643
 QY 167 GTTCCAGATGGGCACTTGGCTATTTCTTCCAGATGTACAGCCCAACATCAGCTCC 226
 Db 142644 GTTCCAGATGGGCACTTGGCTATTTCTTCCAGATGTACAGCCCAACATCAGCTCC 142703
 QY 227 CTGTCTATGAGGACACAGCCGAGCATGACAGCTGTGTTCTTCTTCCGCGCTCT 286
 Db 142704 CTGTCTATGAGGACACAGCCGAGCATGACAGCTGTGTTCTTCTTCCGCGCTCT 142763
 QY 287 GGCCTCTCTTTTGTGGGGCAGCTTTTGGGAGACCCAGAGTGTCTTCTACGTATGGAG 346
 Db 142764 GGCCTCTCTTTTGTGGGGCAGCTTTTGGGAGACCCAGAGTGTCTTCTACGTATGGAG 142823

```

OY 347 CCCGCTGGCAGGTGTGATGACCTTCAAGACAGACCGCGGTGATGCTGCACATGGCT 406
    |||||
Db 142824 CCCGCTGGCAGGTGTGATGACCTTCAAGACAGACCGCGGTGATGCTGCACATGGCT 142883
OY 407 GTGGCGGATCTGATACAGGGCGCTCTTCTGTGGACATGAGCGCTTTGATGCTCAATG 466
    |||||
Db 142884 GTGGCGGATCTGATACAGGGCGCTCTTCTGTGGACATGAGCGCTTTGATGCTCAATG 142943
OY 467 GAACCTGGTCCCGGAGACATCCAGCCCTTTCAGTGGGAGAAAGCCGGGCGCTGTGT 526
    |||||
Db 142944 GAACCTGGTCCCGGAGACATCCAGCCCTTTCAGTGGGAGAAAGCCGGGCGCTGTGT 143003
OY 527 TCTGCACTGCTGTGATCATCTCCACAAGATGAATCATCCCCGGGCTCACTGTCAG 586
    |||||
Db 143004 TCTGCACTGCTGTGATCATCTCCACAAGATGAATCATCCCCGGGCTCACTGTCAG 143063
Y 587 CTCCTGTGATGATCAACAGCCCTTTGAGGTGTGGAGAGAGCGCTGCTCTACAGCCAC 646
    |||||
Db 143064 CTCCTGTGATGATCAACAGCCCTTTGAGGTGTGGAGAGAGCGCTGCTCTACAGCCAC 143123
OY 647 GTGTGCTCAAGAGAGGTGGCTTCTTCAACTGACATGCTCCCTGATCCCGGTGAAAGAC 706
    |||||
Db 143124 GTGTGCTCAAGAGAGGTGGCTTCTTCAACTGACATGCTCCCTGATCCCGGTGAAAGAC 143183
OY 707 CCTCTCCTCACTGATCATCTGTGCACTGTGTCGGGACCCCGGGCGCTGTCCGCTTCC 766
    |||||
Db 143184 CCTCTCCTCACTGATCATCTGTGCACTGTGTCGGGACCCCGGGCGCTGTCCGCTTCC 143243
OY 767 CGAGAAGCGCAAAAGGAGATCTCATGATGACATGCGATGATGTGATGGGAGACATGAG 826
    |||||
Db 143244 CGAGAAGCGCAAAAGGAGATCTCATGATGACATGCGATGATGTGATGGGAGACATGAG 143303
OY 827 CAAAACTCAAAAGAGAGACCAACCTCTATATGTATGATCAGATCTGCCAAAGCCAG 886
    |||||
Db 143304 CAAAACTCAAAAGAGAGACCAACCTCTATATGTATGATCAGATCTGCCAAAGCCAG 143363
OY 887 CTGAGATCTCAAAAGACATCCAGTCTTGGCCAGAGCGCTGCAAGAAAGCTTACCTGCTT 946
    |||||
Db 143364 CTGAGATCTCAAAAGACATCCAGTCTTGGCCAGAGCGCTGCAAGAAAGCTTACCTGCTT 143423
OY 947 GTGCGCTATGAGACCTGCTGAGCGCCCTGTGGCCACAGATCTCCGAATGTATGATGTC 1006
    |||||
Db 143424 GTGCGCTATGAGACCTGCTGAGCGCCCTGTGGCCACAGATCTCCGAATGTATGATGTC 143483
OY 1007 GTGGAGTTGGAAATCTTGGCCCATCTTTCAGACCTGGGTGCATTAACATCAACCGAGCAAG 1066
    |||||
Db 143484 GTGGAGTTGGAAATCTTGGCCCATCTTTCAGACCTGGGTGCATTAACATCAACCGAGCAAG 143543
OY 1067 GGCAATGGGTGACAGGCTTTCACACAAATGGCAGGAGATGCCCTTAATGCTTCCAGGCT 1126
    |||||
Db 143544 GGCAATGGGTGACAGGCTTTCACACAAATGGCAGGAGATGCCCTTAATGCTTCCAGGCT 143603
OY 1127 TGGCGCTGTCTTTCGCTATGAAAGGTTTCGACCTTCAGAAAGCTGTGGCGATGCGC 1186
    |||||
Db 143604 TGGCGCTGTCTTTCGCTATGAAAGGTTTCGACCTTCAGAAAGCTGTGGCGATGCGC 143663
OY 1187 ATGAATTTGCTGGGCTACCGCACGTCAGATCTGAACAAGAACAGAAACCTGTTGCTG 1246
    |||||
Db 143664 ATGAATTTGCTGGGCTACCGCACGTCAGATCTGAACAAGAACAGAAACCTGTTGCTG 143723
OY 1247 GATCTTGTCTACTGAGGCTTCCCTGAGCAATCCATTAAGAGGTTGAAGAGGCTTT 1306
    |||||
Db 143724 GATCTTGTCTACTGAGGCTTCCCTGAGCAATCCATTAAGAGGTTGAAGAGGCTTT 143783
OY 1307 GCTGCCACCTGTGTGACGCTCAGTACATCTTTCGTAATGCTTTCAGGCTTGCCATCAT 1366
    |||||
Db 143784 GCTGCCACCTGTGTGACGCTCAGTACATCTTTCGTAATGCTTTCAGGCTTGCCATCAT 143843
OY 1367 CTCGTAGCCTTAACATCATGTGTGTGGATATCAACATGAGTGTGAGTTGTGTCACACGT 1426
    |||||
Db 143844 CTCGTAGCCTTAACATCATGTGTGTGGATATCAACATGAGTGTGAGTTGTGTCACACGT 143903

```

```

OY 1427 GCTCAAGCAGAAAGGACTTTTGTGTCATGCTGTGTCTAGAAACAGACTGGGAACTT 1486
    |||||
Db 143904 GCTCAAGCAGAAAGGACTTTTGTGTCATGCTGTGTCTAGAAACAGACTGGGAACTT 143963
OY 1487 ATGTAGCAGCAGATCTCCACAGTGAACAGGATATGCTCTTCTTTCTTGTATCTT 1546
    |||||
Db 143964 ATGTAGCAGCAGATCTCCACAGTGAACAGGATATGCTCTTCTTTCTTGTATCTT 144023
OY 1547 CCGTCTGGGAGACTTTCAGAGACTTGTGGCTGAGGCGTATTAAGCAGCAGACAGTA 1606
    |||||
Db 144024 CCGTCTGGGAGACTTTCAGAGACTTGTGGCTGAGGCGTATTAAGCAGCAGACAGTA 144083
OY 1607 TCAGTGAATTTGATTCATTAACCTCCCTGTCACATCTTGCCTAATGGGAAATGATCTT 1666
    |||||
Db 144084 TCAGTGAATTTGATTCATTAACCTCCCTGTCACATCTTGCCTAATGGGAAATGATCTT 144143
OY 1667 TCACCAAGAGCTCACACAGATTTTTCACAGAGATGCT 1703
    |||||
Db 144144 TCACCAAGAGCTCACACAGATTTTTCACAGAGATGCT 144180

RESULT 5
AC138848/c
LOCUS      206943 bp      DNA      linear      HTG 21-JAN-2003
DEFINITION Homo sapiens chromosome 5 clone RP11-1301B21, WORKING DRAFT
ACCESSION  AC138848
VERSION    AC138848.1 GI:27805260
KEYWORDS   HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 206943)
            DOE Joint Genome Institute.
            Sequencing of Human Chromosome 5
            Unpublished
            2 (bases 1 to 206943)
            DOE Joint Genome Institute.
            Direct Submission
            Submitted (21-JAN-2003) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell drive, Walnut Creek, CA 94598, USA
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 2719263
            Center clone name: RCT1-11_1301B21
            -----
            Summary Statistics
            Consensus quality: 202039 bases at least Q40
            Consensus quality: 202804 bases at least Q30
            Consensus quality: 203337 bases at least Q20
            Estimated insert size: 175000; agarose-fp estimation
            Estimated insert size: 206343; sum-of-ctrls estimation
            Quality coverage: 15.28 in Q20 bases; agarose-fp estimation
            Quality coverage: 12.96 in Q20 bases; sum-of-ctrls estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            1      1219: contig of 1219 bp in length
            *      1220      1319: gap of unknown length
            *      1320      2493: contig of 1174 bp in length
            *      2494      2593: gap of unknown length
            *      3665: contig of 1073 bp in length
            *      3667      3766: gap of unknown length

```


TITLE Primer for synthesizing full-length cDNA and use thereof
 JOURNAL Patent: JP 2002017375-A 2689 22-Jan-2002;
 COMMENT HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/2689
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 200253172
 PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
 PI ISHII,
 PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
 SHIMCHI KOJIMA,
 PI TETSUJI OTSUKI, HISASHI KOCA
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/PC
 10, C12P21/02, C12Q1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 Location/Qualifiers
 FT CDS Location/Qualifiers (69) (1226).
 source 1..1979
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 451 a 542 c 474 g 512 t
 ORIGIN
 Query Match 75.7%; Score 1547; DB 6; Length 1979;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 647 CCCCTCCCTCAACCTGCATATCGTGCACCTGGTCCGGAGCCCCGGGCGGTTCGGTTC 706
 QY 766 CCGAGAACGACAAAAGGAGATCTCATGATTTGACATCCGATTTGATGGGACGATGA 825
 Db 707 CCGAGAACGACAAAAGGAGATCTCATGATTTGACATCCGATTTGATGGGACGATGA 766
 QY 826 GCAAAAACGCAAGAGAGAGACCAACCTACTATGTGATGACAGTATCTGCCAAGCA 885
 Db 767 GCAAAAACGCAAGAGAGAGACCAACCTACTATGTGATGACAGTATCTGCCAAGCA 826
 QY 886 GCTGGAGATCTCAAGACATCCAGTCTTGGCCAAAGGCGCTGCGAGAACGCTACTGCT 945
 Db 827 GCTGGAGATCTCAAGACATCCAGTCTTGGCCAAAGGCGCTGCGAGAACGCTACTGCT 886
 QY 946 TGTGCGCTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTTCCCGATGTATGAAAT 1005
 Db 887 TGTGCGCTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTTCCCGATGTATGAAAT 946
 QY 1006 CGTGGGATTTGGAATTTTGGCCCATCTTACAGACTGGTGGATTAACATCACCAGGCAA 1065
 Db 947 CGTGGGATTTGGAATTTTGGCCCATCTTACAGACTGGTGGATTAACATCACCAGGCAA 1006
 QY 1066 GGGCATGGGTGAGACCAAGCTTTCACACAAATGCGAGAGTGCCTTAATGTCCAGGC 1125
 Db 1007 GGGCATGGGTGAGACCAAGCTTTCACACAAATGCGAGAGTGCCTTAATGTCCAGGC 1066
 QY 1126 TTGGCGCTGTCTTTGCCCCTATGAAAAGTTTCTGACCTTCAGAAAAGCCTGTGGCATGC 1185
 Db 1067 TTGGCGCTGTCTTTGCCCCTATGAAAAGTTTCTGACCTTCAGAAAAGCCTGTGGCATGC 1126
 QY 1186 CATGAATTTGCTGGGTACCGGCAGCTCAGATCTGAAACAAGAACAGAAACCTGTGCT 1245
 Db 1127 CATGAATTTGCTGGGTACCGGCAGCTCAGATCTGAAACAAGAACAGAAACCTGTGCT 1186
 QY 1246 GGAATCTTCTGTACTCTGAGTCTGCCAGAGAAATCCACTAAGAGGTTGAGAAAGCTT 1305
 Db 1187 GGAATCTTCTGTACTCTGAGTCTGCCAGAGAAATCCACTAAGAGGTTGAGAAAGCTT 1246
 QY 1306 TGTGCGCACTGGTGTGACCTTCAGCTTCGTAATGCTTCTGAGCCTTGCCTACA 1365
 Db 1247 TGTGCGCACTGGTGTGACCTTCAGCTTCGTAATGCTTCTGAGCCTTGCCTACA 1306
 QY 1366 TCTCTGAGGCTTAACATACATGCTGAGATACATGAGTGTGAGTGTGCTGACACAG 1425
 Db 1307 TCTCTGAGGCTTAACATACATGCTGAGATACATGAGTGTGAGTGTGCTGACACAG 1366
 QY 1426 TGTCTAAGCAGAGAGACTTTTGTCTCATGCTTGTCTGAGAAAACAGACTGGGAACT 1485
 Db 1367 TGTCTAAGCAGAGAGACTTTTGTCTCATGCTTGTCTGAGAAAACAGACTGGGAACT 1426
 QY 1486 TATGTGACGACATCCACACAGTGAAGAGGATATTCCTCTCTCTTCTTCTTGATCT 1545
 Db 1427 TATGTGACGACATCCACACAGTGAAGAGGATATTCCTCTCTCTTCTTCTTGATCT 1486
 QY 1546 TCTCTGTGAGGAGACTTTCAGAGACTTGTGGCGTGGAGGCTTAATTAACACAGACACT 1605
 Db 1487 TCTCTGTGAGGAGACTTTCAGAGACTTGTGGCGTGGAGGCTTAATTAACACAGACACT 1546
 QY 1606 ATCAGTGAATTTGATCATTAACCTCCCTGTCCACATCTTGGCCATGGGAGATGATCT 1665
 Db 1547 ATCAGTGAATTTGATCATTAACCTCCCTGTCCACATCTTGGCCATGGGAGATGATCT 1606
 QY 1666 TTACCAAAAGAGCTCACAGCATTTTTCACAGAGATGC 1703
 Db 1607 TTACCAAAAGAGCTCACAGCATTTTTCACAGAGATGC 1644
 RESULT 7
 AK074746 1979 bp mRNA linear PRI 03-SEP-2002
 LOCUS Homo sapiens cDNA FLJ90265 fls. clone NTBM4001325, weakly similar
 DEFINITION to CHONDROITIN 6-SULFOTRANSFERASE (EC 2.8.2.17).
 ACCESSION AK074746

```

VERSION      AK074746.1 GI:22760388
KEYWORDS     oligo capping; fls (full insert sequence).
SOURCE       Homo sapiens (human)
ORGANISM     Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
              Eukaryota; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1
AUTHORS      Isogai,T., Ota,T., Nishikawa,T., Hayashi,K., Otsuki,T.,
              Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S.,
              Kawai-Hio,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y.,
              Kojima,S., Nagahara,K., Masuno,Y., Ono,T., Okano,K., Yoshikawa,Y.,
              Aotsuka,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and
              Niimiyu,K.
TITLE        NEDO human cDNA sequencing project
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 1979)
AUTHORS      Isogai,T. and Otsuki,T.
TITLE        Direct Submission
JOURNAL      Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute,
              Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
              (E-mail:genomices@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT      NEDO human cDNA sequencing project supported by Ministry of
              Economy, Trade and Industry of Japan; cDNA full insert sequencing;
              Research Association for Biotechnology; cDNA library construction;
              Institute of Medical Science, University of Tokyo, Laboratory of
              Genome Structure, Human Genome Center; cDNA 5'-6' end one pass
              sequencing and clone selection; Helix Research Institute (supported
              by Japan Key Technology Center etc.).
FEATURES     Location/Qualifiers
              1..1979
                source          /organism="Homo sapiens"
                               /mol_type="mRNA"
                               /db_xref="taxon:9606"
                               /clone="NT2RM4.001325"
                               /cell_line="NT2"
                               /cell_type="teratocarcinoma"
                               /clone_lib="NT2RM4"
                               /note="cloning vector: PME18SFL3-mRNA from uninduced NT2
                               neuronal precursor cells"
                               69..1229
                               /note="unlabeled protein product"
                               /codon_start=1
                               /protein_id="BAC11177.1"
                               /db_xref="GI:22760389"
                               /translation="MLPDKKLLFLVSOMATLAFEHYSHNISLSKMADEBWH
                               VLISSMRSSSFVGLFGHDPVFLMEFMAHVMFPGSTAMLMHARLDIRAF
                               LCDMSYFDATMEBPGRSSSLPOMENRALSAPACDITFQDELIPRRLCSQP
                               FEVEKACRSYSHVVLKEVFENLQSLYPLKDPISLHATVHLVRDPAVERSERK
                               GDLMDINRIYMGHQLKKEKEDPYVMQVTCOSLEIYTIQSLTPRALDERLYLVY
                               EDLARAIVAQTSRMFEVGLFELPHLOTVHNITRGGMGDHAFHTVARADALNSQAW
                               RWSLPYKVRRLKACGDANLGLYRHVRSEQRNLIDLSTWYVPDQIH"
              451 a             542 c             474 g             512 t
              RIGIN
              ASE COUNT
              Query Match      75.7%; Score 1547; DB 9; Length 1979;
              Best Local Similarity 99.9%; Pred. No. 0;
              Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 346 GCCGCCGCGCAGCGTGGATGACCTTCAAGCAGACGCGCTGGATGCTGCACATGGC 405
    |||||||
Db 287 GCCCGCCCGCAGCGTGGATGACCTTCAAGCAGACGCGCTGGATGCTGCACATGGC 346
Oy 406 TGTGCGGAGATCTGATACGCGCGCTCTTGTGGACATGAGCGCTTTGATGGCTACAT 465
    |||||||
Db 347 TGTGCGGAGATCTGATACGCGCGCTCTTGTGGACATGAGCGCTTTGATGGCTACAT 406
Oy 466 GGAACCTGTGTCCTCCGAGACAGTCAGCTCTTTCAGTGGAGAAACAGCGCGCTGTG 525
    |||||||
Db 407 GGAACCTGTGTCCTCCGAGACAGTCAGCTCTTTCAGTGGAGAAACAGCGCGCTGTG 466
Oy 526 TTTCGACACCTGCGGTGATACATCCACAAAGATGAATATCCCGCGGCTCACTGAG 585
    |||||||
Db 467 TTTCGACACCTGCGGTGATACATCCACAAAGATGAATATCCCGCGGCTCACTGAG 526
Oy 586 GCTCCTGTGACATACAGCCCTTGTAGGTGGTGGAGAGCGCTGCGCTCTACAGCA 645
    |||||||
Db 527 GCTCCTGTGACATACAGCCCTTGTAGGTGGTGGAGAGCGCTGCGCTCTACAGCA 586
Oy 646 CTGTGCTCTCAAGAGGTGCGCTTCTCAACCTGCACTCCCTTACCCGCTGTGAAGA 705
    |||||||
Db 587 CTGTGCTCTCAAGAGGTGCGCTTCTCAACCTGCACTCCCTTACCCGCTGTGAAGA 646
Oy 706 CCCCTCCCTCAACCTGCACTATCGACACCTGCTCCGAGACCCCGCGGCTTCCGTTG 765
    |||||||
Db 647 CCCCTCCCTCAACCTGCACTATCGACACCTGCTCCGAGACCCCGCGGCTTCCGTTG 706
Oy 766 CCGAGAACGCAAAAGGAGATCTCATGATGACAGTGCATGTGTGATGGGCGAGATGA 825
    |||||||
Db 707 CCGAGAACGCAAAAGGAGATCTCATGATGACAGTGCATGTGTGATGGGCGAGATGA 766
Oy 826 GCAAAACTCAAGAAGAGAGACCAACCTACTATGTGATGAGTCACTGCTCCAAAGCA 885
    |||||||
Db 767 GCAAAACTCAAGAAGAGAGACCAACCTACTATGTGATGAGTCACTGCTCCAAAGCA 826
Oy 886 GCTGAGATCTACAAAGACATCCAGTCTTCCCAAGGCGCTGGAGAAAGCTACCTGCT 945
    |||||||
Db 827 GCTGAGATCTACAAAGACATCCAGTCTTCCCAAGGCGCTGGAGAAAGCTACCTGCT 886
Oy 946 TGTGCGCTATGAGAGACCTGCTGAGCCCTGTGGCCAGACTTCCGAATGTATGAATT 1005
    |||||||
Db 887 TGTGCGCTATGAGAGACCTGCTGAGCCCTGTGGCCAGACTTCCGAATGTATGAATT 946
Oy 1006 CGTGGGATTTGAATTTTGGCCCATCTTTCAGACCTGGGTGATTAACATCACCCAGGCA 1065
    |||||||
Db 947 CGTGGGATTTGAATTTTGGCCCATCTTTCAGACCTGGGTGATTAACATCACCCAGGCA 1006
Oy 1066 GGGCATGGGACACAGCTTTCACACAAATGGCAGGAGATGCCCTTAATGTCTCCAGGC 1125
    |||||||
Db 1007 GGGCATGGGACACAGCTTTCACACAAATGGCAGGAGATGCCCTTAATGTCTCCAGGC 1066
Oy 1126 TTGGCGCTGTGCTTTGGCCATATGAAAAGTTTCTCGACTTCAGAAAGCTGTGGCATGC 1185
    |||||||
Db 1067 TTGGCGCTGTGCTTTGGCCATATGAAAAGTTTCTCGACTTCAGAAAGCTGTGGCATGC 1126
Oy 1186 CATGAATTTGCTGGGCTACCGCCACGCTCAGATCTGAAACAGAAAGAAACCTGTGCT 1245
    |||||||
Db 1127 CATGAATTTGCTGGGCTACCGCCACGCTCAGATCTGAAACAGAAAGAAACCTGTGCT 1186
Oy 1246 GGATCTCTGCTCTACCTGAGATGCTCCCTGAGCAATGCACATGAGAGGTTGAGAAGCTT 1305
    |||||||
Db 1187 GGATCTCTGCTCTACCTGAGATGCTCCCTGAGCAATGCACATGAGAGGTTGAGAAGCTT 1246
Oy 1306 TGTGCGACCTGTGTGAGCTCAGTCACTTCTCTGAATGCTTCTGAGCCTTGCCTACA 1365
    |||||||
Db 1247 TGTGCGACCTGTGTGAGCTCAGTCACTTCTCTGAATGCTTCTGAGCCTTGCCTACA 1306
Oy 1366 TCTCTGAGCCTTAACTACATGTCTGTGGGATATCACTAGATGTGATGTGTGTCAACAG 1425
    |||||||
Db 1307 TCTCTGAGCCTTAACTACATGTCTGTGGGATATCACTAGATGTGATGTGTGTCAACAG 1366

```



```

QY 1426 TCCTCAAGCAGAGACTTTGTGTCATGCTTGTGTCTAGAAAAACAGTGGGAACT 1485
DB 1367 TCCTCAAGCAGAGACTTTGTGTCATGCTTGTGTCTAGAAAAACAGTGGGAACT 1426
QY 1486 TATGTAGAGCAGACATCCACAGTGAAGAGGATTTGCTCTCTCTCTCTCTCTCTCT 1545
DB 1427 TATGTAGAGCAGACATCCACAGTGAAGAGGATTTGCTCTCTCTCTCTCTCTCTCT 1486
QY 1546 TCCTGTCTGGGAGACTTCAGAGACTTTGTGGCTGGAGGCTATTAAACAGCAGACAGT 1605
DB 1487 TCCTGTCTGGGAGACTTCAGAGACTTTGTGGCTGGAGGCTATTAAACAGCAGACAGT 1546
QY 1606 ATCAGTGAATGATTCATTAACCTCCCTGTCACATCTTGGCCAAATGGGAAATGATCT 1665
DB 1547 ATCAGTGAATGATTCATTAACCTCCCTGTCACATCTTGGCCAAATGGGAAATGATCT 1606
QY 1666 TTCACCAAGAGCTCACAGACTTTTCCACAGAGATGC 1703
DB 1607 TTCACCAAGAGCTCACAGACTTTTCCACAGAGATGC 1644

RESULT 8
AR203335 AR203335 2032 bp DNA linear PART 20-JUN-2002
LOCUS Sequence 2 from patent US 6365365.
DEFINITION AR203335
ACCESSION AR203335 GI:21499698
VERSION AR203335.1
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2032)
AUTHORS Bistrup,A., Rosen,S.D., Tangemann,K. and Hemmerich,S.
TITLE Method of determining whether an agent modulates glycosyl
sulfoltransferase-3
JOURNAL Patent: US 6365365-A 2 02-APR-2002;
FEATURES
source Location/Qualifiers
BASE COUNT 468 a 569 c 490 g 505 t
ORIGIN

Query Match 75.7%; Score 1547; DB 6; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 526 TTCTGCACCTGCGCTGTGACATCATCCACAGATGAATATCCCGGGGCTCCTACAG 585
DB 580 TTCTGCACCTGCGCTGTGACATCATCCACAGATGAATATCCCGGGGCTCCTACAG 639
QY 586 GCTCTCTGTGACATCAACAGCCCTTTAGAGTGTGGAGAAAGGCTGCCCTCTACAGCA 645
DB 640 GCTCTCTGTGACATCAACAGCCCTTTAGAGTGTGGAGAAAGGCTGCCCTCTACAGCA 699
QY 646 CGTGTGTCTCAAGAGAGTGGCTTTCTCAACCTGCAGTCCCTCTACCCCTGCTGAAGA 705
DB 700 CGTGTGTCTCAAGAGAGTGGCTTTCTCAACCTGCAGTCCCTCTACCCCTGCTGAAGA 759
QY 706 CCCCTCTCTCAACCTGCATATGCTGACCTGTCCGGAGCCCGGGCCGTGTTCCGTTTC 765
DB 760 CCCCTCTCTCAACCTGCATATGCTGACCTGTCCGGAGCCCGGGCCGTGTTCCGTTTC 819
QY 766 CCGAGAACCCACAAAGGAGATCTCATGATGATGACATCCCATTTGATGGGGAGAGATA 825
DB 820 CCGAGAACCCACAAAGGAGATCTCATGATGATGACATCCCATTTGATGGGGAGAGATA 879
QY 826 GCNAAAACCTCAAGAGAGAGACCAACCTCTATGTGATGACAGTCTGCCAAAGCCA 885
DB 880 GCNAAAACCTCAAGAGAGAGACCAACCTCTATGTGATGACAGTCTGCCAAAGCCA 939
QY 886 GCTGTGAGATCTACAGAACCATCCAGTCTTGGCCCAAGGCCCTGACAGAAAGCTACCTGCT 945
DB 940 GCTGTGAGATCTACAGAACCATCCAGTCTTGGCCCAAGGCCCTGACAGAAAGCTACCTGCT 999
QY 946 TGTGGCTATGAGAGAGCTGGCTGAGCCCTGTGGGCCACAGACTTCCGAAATGATTAAT 1005
DB 1000 TGTGGCTATGAGAGAGCTGGCTGAGCCCTGTGGGCCACAGACTTCCGAAATGATTAAT 1059
QY 1006 CGTGGGATTTGGAATTTCTCCCATCTTGTGAGACTGGGTGATCAACATCAACCGAGCAA 1065
DB 1060 CGTGGGATTTGGAATTTCTCCCATCTTGTGAGACTGGGTGATCAACATCAACCGAGCAA 1119
QY 1066 GGGCATGGGTGACACACGCTTTCCACACAATAATGCGAGATGCCCTTAATGTCTCCAGGC 1125
DB 1120 GGGCATGGGTGACACACGCTTTCCACACAATAATGCGAGATGCCCTTAATGTCTCCAGGC 1179
QY 1126 TTGGGCTGGTCTTTCCTTCCATGAAAAGTTCTGACATTCAGAAAAGCCGTGGGAGATC 1185
DB 1180 TTGGGCTGGTCTTTCCTTCCATGAAAAGTTCTGACATTCAGAAAAGCCGTGGGAGATC 1239
QY 1186 CATGAATTTGTGGGCTTACCGCCACGCTCAGATCTGAACAAGAACAGAGAACTGTGGCT 1245
DB 1240 CATGAATTTGTGGGCTTACCGCCACGCTCAGATCTGAACAAGAACAGAGAACTGTGGCT 1299
QY 1246 GGATCTTGTGTCTACCTGAGACTGTCCCTGAGCAATTCACATAAGAGGTTGAGAGGCTT 1305
DB 1300 GGATCTTGTGTCTACCTGAGACTGTCCCTGAGCAATTCACATAAGAGGTTGAGAGGCTT 1359
QY 1306 TGTGTGCACCTGGTGTACAGCCAGTCAGTCTTCTGATAGTGTGTGAGCCTTGCTTACA 1365
DB 1360 TGTGTGCACCTGGTGTACAGCCAGTCAGTCTTCTGATAGTGTGTGAGCCTTGCTTACA 1419
QY 1366 TCTCTGAGCCTTAACATACATGTCTGTGGGTATACACTAGTGTGATGTGTCTCACAG 1425
DB 1420 TCTCTGAGCCTTAACATACATGTCTGTGGGTATACACTAGTGTGATGTGTCTCACAG 1479
QY 1426 TCCTCAAGCAGAGACTTTGTGTGTCATGCTTGTGTCTAGAAAACAGTGGGAACT 1485
DB 1480 TCCTCAAGCAGAGACTTTGTGTGTCATGCTTGTGTCTAGAAAACAGTGGGAACT 1539
QY 1486 TATGTAGAGCAGACATCCACAGTGAAGAGGATTTGCTCTCTCTCTCTCTCTCTCT 1545
DB 1540 TATGTAGAGCAGACATCCACAGTGAAGAGGATTTGCTCTCTCTCTCTCTCTCTCT 1599
QY 1546 TCCTGTCTGGGAGACTTCAGAGACTTTGTGGCTGGAGGCTATTAAACAGCAGACAGT 1605
DB 1600 TCCTGTCTGGGAGACTTCAGAGACTTTGTGGCTGGAGGCTATTAAACAGCAGACAGT 1659

```


Db	1540	TATGTGACACACATCCACCAGTGAAGAAGGGGATATGCTTCCTCCTTTTCCTGANTCT	1599
Oy	1546	TCCTGTCTGGGCACACTTCACAGACTTTGTGTGCGCTTGAGGCCCTATTAAAGCACAGACAGT	1605
Db	1600	TCCTGTCTGGGCACACTTCACAGACTTTGTGTGCGCTTGAGGCCCTATTAAAGCACAGACAGT	1659
Oy	1606	ATCATGTGAATTGATCCATAAACCCTCCCTGCCACATCTTCGCCAATGGGGAATGAGTCT	1665
Db	1660	ATCATGTGAATTGATCCATAAACCCTCCCTGCCACATCTTCGCCAATGGGGAATGAGTCT	1719
Db	1720	TTCAACAAGAGCTCACACAGCATTTTTCACAGAGATGC	1757
RESULT 10			
AF131235		2032 bp	mRNA linear PRI 10-AUG-1999
DEFINITION	Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA,		
complete cds.			
AF131235			
AF131235.1	GI:4927113		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Euthelia; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 2032)		
JOURNAL	Bistrup,A., Bhakta,S., Lee,J.K., Belov,Y.Y., Gunn,M.D., Zuo,F.R.,		
MEDLINE	Huang,C.C., Kannagi,R., Rosen,S.D. and Hemmerich,S.		
PUBMED	Sulfotransferases of two specificities function in the		
REFERENCE	reconstitution of high endothelial cell ligands for L-selectin		
AUTHORS	J. Cell Biol. 145 (4), 899-910 (1999)		
TITLE	2 (bases 1 to 2032)		
JOURNAL	Bistrup,A., Tangemann,K., Bhakta,S., Lee,J.-K., Belov,Y.Y.,		
MEDLINE	Gunn,M.D., Zuo,F.-R., Huang,C.-C., Kannagi,R., Rosen,S.D. and		
PUBMED	Hemmerich,S.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (25-FEB-1999) Respiratory Diseases, Roche Bioscience,		
TITLE	3401 Hillview Avenue, Palo Alto, CA 94304, USA		
JOURNAL	Location/Qualifiers		
FEATURES	1..2032		
SOURCE	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/cell_type="high endothelial"		
	/tissue_type="tonsil"		
CDS	182..1342		
	/function="biosynthesis of endothelial ligands for		
	L-selectin		
	/note="HEC-GlCNAC-6-S'T"		
	/codon_start=1		
	/product="N-acetylglucosamine 6-O-sulfotransferase"		
	/protein_id="AAD33015.1"		
	/db_xref="GI:4927114"		
	/translation="MLPKMKLLFLVSOAILLFFMYSHNISLSKKAOPERMH		
	VLVTVSWSSSEFVGQLEGFQHPDEYELAMEPAHWMTFKOSTAMLMAYDLIRAVF		
	LCIDSVDAWMEPCGPROSSLDIOWENSRALCSAPACIIIPODETIIPAHRLCSOOP		
	GEYEMKACRSYSHVLKEVFEPNIOSLYPLKDPSILHLHYDRPVRPRSERERK		
	GDALMDRIYMGQHEOKLKREDPDITYMOYCQSOLEITRTIOSLPALDERLTIVTY		
	EDLARAEVAQTSTKTEFVGLFEPLPHLOTWHNTITRGMGDHAHTVARALNVASQM		
	RMSIPLYEKVSRLOKACDAMNNLLGYRVRSEOEORNLIDLTLSTWVPEOIH"		
BASE COUNT	468 a 569 c 490 g 505 t		
ORIGIN			
Query Match	75.7%; Score 1547; DB 9; Length 2032;		
Best Local Similarity	99.9%; Pred. No. 0;		
Matches 1597; Conservative	0; Mismatches 1; Indels 0; Gaps 0;		
Oy	106	CAAGCTCTTCACATTGACCAACAANGCNCACCTTAATAAATAAGCTCCTGCTGTTTCT	165

Db	160	AAAGTCTTCCACTTCACACACATGCTACTGCTTAAAAAATGAAAGCTCCGTGCTTCT	219
OY	166	GGTTTCCAGATGGCCATCTTGCTGTATCTTCCATATGTACAGCCACAACATCAGCTC	225
Db	220	GGTTTCCAGATGGCCATCTTGCTGTATCTTCCATATGTACAGCCACAACATCAGCTC	279
OY	226	CCTGTCTATGAAGGCACAGCCCGAGGGCATGACGTGCTGTCTCTTCTGTGGCGCTC	285
Db	280	CCTGTCTATGAAGGCACAGCCCGAGGGCATGACGTGCTGTCTCTTCTGTGGCGCTC	339
OY	286	TGGCTCTTCTTTTGTGGGGCAGCTTTTGGGACAGACCAGATGTTTCTACTATGGA	345
Db	340	TGGCTCTTCTTTTGTGGGGCAGCTTTTGGGACAGACCAGATGTTTCTACTATGGA	399
OY	346	GCCCGCCTGGCAGCTGTGTGATGACCTTCAACGACAGCACCGCCTGGATGCTGCACATGGC	405
Db	400	GCCCGCCTGGCAGCTGTGTGATGACCTTCAACGACAGCACCGCCTGGATGCTGCACATGGC	459
OY	406	TGTGCGGGATCTGATACGGGCCGTCTTGTGTGGACATGAGCGCTTTGATGCTTACAT	465
Db	460	TGTGCGGGATCTGATACGGGCCGTCTTGTGTGGACATGAGCGCTTTGATGCTTACAT	519
OY	466	GGAACCTGTGTCGGGAGACAGTCCAGCCTTTTACAGGGGAGAACACCGGGCCCTGTG	525
Db	520	GGAACCTGTGTCGGGAGACAGTCCAGCCTTTTACAGGGGAGAACACCGGGCCCTGTG	579
OY	526	TTTGTGCACCTGCTGTGACATCATCCACAAGATGAATAATCATCCCGCGGCTCACTGCAG	585
Db	580	TTTGTGCACCTGCTGTGACATCATCCACAAGATGAATAATCATCCCGCGGCTCACTGCAG	639
OY	586	GCTCCTGTGCAGTCAACAGCCCTTTGAGGTGTGAGAGAGCGCTGCGCTCCTTACAGCCA	645
Db	640	GCTCCTGTGCAGTCAACAGCCCTTTGAGGTGTGAGAGAGCGCTGCGCTCCTTACAGCCA	699
OY	646	CGTGTGTCTAAGGGAGTGGCGCTTCTTAACCTGAGTGCCTTCAACCGGCTGTGAAAGA	705
Db	700	CGTGTGTCTAAGGGAGTGGCGCTTCTTAACCTGAGTGCCTTCAACCGGCTGTGAAAGA	759
OY	706	CCCTCTCCCTCAACCTGTATCTGTGCACCTGGTCCGGGACCCCGGGCGTGTCCGTTTC	765
Db	760	CCCTCTCCCTCAACCTGTATCTGTGCACCTGGTCCGGGACCCCGGGCGTGTCCGTTTC	819
OY	766	CCGAGAACGCACAAAGGAGATCTTCATGTATGTACAGTGGCATTTGTGATGGGGCAGCATGA	825
Db	820	CCGAGAACGCACAAAGGAGATCTTCATGTATGTACAGTGGCATTTGTGATGGGGCAGCATGA	879
OY	826	GCAAAAACCTCAGAGAGAGACCAACCTTACTATGTGATGTGAGAGTCACTGCCAAAGCCA	885
Db	880	GCAAAAACCTCAGAGAGAGACCAACCTTACTATGTGATGTGAGAGTCACTGCCAAAGCCA	939
OY	886	GCTGAGAGTCTACAAGACCATCCAGTCCCTTGGCCCAAGGCGCTGCAGAAAGCCTACCTGCT	945
Db	940	GCTGAGAGTCTACAAGACCATCCAGTCCCTTGGCCCAAGGCGCTGCAGAAAGCCTACCTGCT	999
OY	946	TGTGCGCTATAGAGACCTGTGCTCGAGCCCTGTGTGGCCCAAGCTTCCGAAATGTATGAATT	10059
Db	1000	TGTGCGCTATAGAGACCTGTGCTCGAGCCCTGTGTGGCCCAAGCTTCCGAAATGTATGAATT	10599
OY	1006	CGTGGGATTTGAAATTTCTTGCCCAATCTTCAACACCTGGGTGATTAACATCACCCGAGCAA	10655
Db	1060	CGTGGGATTTGAAATTTCTTGCCCAATCTTCAACACCTGGGTGATTAACATCACCCGAGCAA	11199
OY	1066	GGGCGATGGGTGACCAAGCTTTTCCACACAAATGCCAGGGATCCCTTAATGTCTCCAGGC	11259
Db	1120	GGGCGATGGGTGACCAAGCTTTTCCACACAAATGCCAGGGATCCCTTAATGTCTCCAGGC	11799
OY	1126	TTGGCGCTGTCTTTTGCCCTATGAAAAGTTTCTGACTTCAGAAAAGCTGTGGCGATGC	11855
Db	1180	TTGGCGCTGTCTTTTGCCCTATGAAAAGTTTCTGACTTCAGAAAAGCTGTGGCGATGC	12399
OY	1186	CATGAATTTGTGGGGCTACCGCCGACGTCAGATCTGGAACAACAAGAAACCTGTGGCT	12455
Db	1240	CATGAATTTGTGGGGCTACCGCCGACGTCAGATCTGGAACAACAAGAAACCTGTGGCT	12999

QY 1246 GGATCTTCTGTCTACTGAGCTGCTGAGCAATCCACTAAGAGGTTGAGAAGCTT 1305
 Db 1300 GGATCTTCTGTCTACTGAGCTGCTGAGCAATCCACTAAGAGGTTGAGAAGCTT 1359
 QY 1306 TGTGACCACTGTGTGACGCTCAGTCACTTTCTGTGAATGCTTGTGAGCCTTGCTACA 1365
 Db 1360 TGTGACCACTGTGTGACGCTCAGTCACTTTCTGTGAATGCTTGTGAGCCTTGCTACA 1419
 QY 1366 TCTGTGACCTTACTCATGCTGTGTGATCATACAGTGTGAGTGTGTCTCCACAG 1425
 Db 1420 TCTGTGACCTTACTCATGCTGTGTGATCATACAGTGTGAGTGTGTCTCCACAG 1479
 QY 1426 TGTCTAAGCAAGAGACTTTTGTCTCATGCTGTGTGTAGAAAACAGACGAGGAACT 1485
 Db 1480 TGTCTAAGCAAGAGACTTTTGTCTCATGCTGTGTGTGTAGAAAACAGACGAGGAACT 1539
 QY 1486 TATGTGACAGACATCCACAGTGAACAGAGGTATGCTTCTTCTTGTGTATCT 1545
 Db 1540 TATGTGACAGACATCCACAGTGAACAGAGGTATGCTTCTTCTTGTGTATCT 1599
 QY 1546 TCTGTGTGGGACAGTTCAGAGACTTGTGGCTGGAGGCTATTAAGCAGACAGCT 1605
 Db 1600 TCTGTGTGGGACAGTTCAGAGACTTGTGGCTGGAGGCTATTAAGCAGACAGCT 1659
 QY 1606 ATCAGTGAATGTATCCATAAACCTCCCTGTCCACATCTTGTCCCAATGGGAAATGATCT 1665
 Db 1660 ATCAGTGAATGTATCCATAAACCTCCCTGTCCACATCTTGTCCCAATGGGAAATGATCT 1719
 QY 1666 TTCACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 1703
 Db 1720 TTCACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 1757

RESULT 11
 AF149783 1333 bp mRNA linear PRI 02-JUL-2001
 LOCUS AF149783 Homo sapiens L-selectin ligand sulfotransferase mRNA, complete cds.
 DEFINITION AF149783
 VERSION AF149783.1 GI:13897503
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 1333)
 Yeh,J.C., Hirakawa,N., Petryniak,B., Nakayama,J., Elles,L.G.,
 Raduka,D., Hindsgraul,O., Marth,J.D., Lowe,J.B. and Fukuda,M.
 Novel sulfated lymphocyte homing receptors and their control by a
 corel extension beta 1,3-N-acetylglucosaminyltransferase
 Cell 105 (7), 957-969 (2001)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 SOURCE

1. 1333
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 93.1253
 /note="LSST"
 /codon_start=1
 /product="L-selectin ligand sulfotransferase"
 /protein_id="AAK48417.1"
 /db_xref="GI:13897504"
 /translation="MLPKMKLLPLVSVMAIALFFHMYSHNISLSMKADPERMH
 VLIVSSRSSSEFVGLQHPDVEFLMEAMHWMFECSTAMLMIAVADLIRAF
 LCMDSVDAEMEPKRSRSLFQWENRALSAPACDIIDDEIIPARHRLCSQOP
 FEVEKACRSYSHVLKEVREFNLSLYPLTKDPSLMILHIVLRDRAVERSRERIK

BASE COUNT 286 a 393 c 336 g 318 t
 ORIGIN
 Query Match 61.8%; Score 1262; DB 9; Length 1333;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 GDLMDISRTVMGHOEKLKEDOPYYVMQVICSOLIEYKTIQSLPKAIOERLLVRY
 EDLAPAPVQOTSRMYEFGLEPLHLOTNWINTRGKMGSHAFETNRDLNYSQAM
 RWSLPIYEVSHLOACGDMANMLILRYHRHSRQBRNLLDLSTWTVPEQJH"
 107 AAGCTCTTCACTTACAGCAATGCTACTGCTTAAAAAATGAAAGCTCTGCTTTGTG 166
 72 AAGCTCTTCACTTACAGCAATGCTACTGCTTAAAAAATGAAAGCTCTGCTTTGTG 131
 167 GTTCCCGATGGGCACTTGGCTATCTTCCAGATGATGAGCAACAATGAGCTCC 226
 132 GTTCCCGATGGGCACTTGGCTATCTTCCAGATGATGAGCAACAATGAGCTCC 191
 227 CTGTCTATGAAGGACAGCCGAGGAGCATGACAGTGTGTGCTTCTTCTGCGCTCT 286
 192 CTGTCTATGAAGGACAGCCGAGGAGCATGACAGTGTGTGCTTCTTCTGCGCTCT 251
 287 GGTCTTTCTTTTGTGTGGGACAGCTTTTGTGGACAGCACCAGATGTTTCTACTGATGAG 346
 252 GGTCTTTCTTTTGTGTGGGACAGCTTTTGTGGACAGCACCAGATGTTTCTACTGATGAG 311
 347 CCCGCTGGACAGTGTGATGACCTTCAAGAGACAGCAGCCGCTGATGCTGACATGCT 406
 312 CCCGCTGGACAGTGTGATGACCTTCAAGAGACAGCAGCCGCTGATGCTGACATGCT 371
 407 GTGGGGATCTGATACGGGCGCTTCTTGTGCGACATGAGGCTTGTGATGCTTACATG 466
 372 GTGGGGATCTGATACGGGCGCTTCTTGTGCGACATGAGGCTTGTGATGCTTACATG 431
 467 GAACCTGTCCTCCCGAGAGACATCCAGCTCTTTCAGTGGAGAACAGCCGGCGCTGTGT 526
 432 GAACCTGTCCTCCCGAGAGACATCCAGCTCTTTCAGTGGAGAACAGCCGGCGCTGTGT 491
 527 TCTGACACCTGGCTGTGACATCATCCACAGATGAATCATCCCGGCGCTGACATGAGG 586
 492 TCTGACACCTGGCTGTGACATCATCCACAGATGAATCATCCCGGCGCTGACATGAGG 551
 587 CTCTGTGACATGACACAGCCCTTGTGAGTGTGTGAGAGGCTGCGCTCTTACACAGCC 646
 552 CTCTGTGACATGACACAGCCCTTGTGAGTGTGTGAGAGGCTGCGCTCTTACACAGCC 611
 647 GTGTGTCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTCTACCCCTGCTGAAGAC 706
 612 GTGTGTCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTCTACCCCTGCTGAAGAC 671
 707 CCCCTCCCAACCTGCAATGCTGACAGCTGTGCGGAGACCCCGGCGCTGTCCGCTCC 766
 672 CCCCTCCCAACCTGCAATGCTGACAGCTGTGCGGAGACCCCGGCGCTGTCCGCTCC 731
 767 CGAGAACGACAAAAGGAGATCTCATGATGATGACAGTCCATTTGTATGGGCGAGCATGAG 826
 732 CGAGAACGACAAAAGGAGATCTCATGATGATGACAGTCCATTTGTATGGGCGAGCATGAG 791
 827 CAAAAAATCAAGAGAGAGACCAACCTACTATGTGATGACAGTGTATGTCGCAAGCCAG 886
 792 CAAAAAATCAAGAGAGAGACCAACCTACTATGTGATGACAGTGTATGTCGCAAGCCAG 851
 887 CTGAGATCTCAAGAGACATCCAGTCTTGTGCGCAAGGCGCTGAGGAACGCTACTGCTT 946
 852 CTGAGATCTCAAGAGACATCCAGTCTTGTGCGCAAGGCGCTGAGGAACGCTACTGCTT 911
 947 GTGGCATGTAGAGAGCTGGCTGAGGCGCTGTGGCGCCAGATCCCGAATGATGAATTC 1006
 912 GTGGCATGTAGAGAGCTGGCTGAGGCGCTGTGGCGCCAGATCCCGAATGATGAATTC 971
 1007 GTGGATTTGAATCTTGTCCCATCTTCAAGCTGTGGTGTATTAACATCACCCGAGGCAAG 1066
 972 GTGGATTTGAATCTTGTCCCATCTTCAAGCTGTGGTGTATTAACATCACCCGAGGCAAG 1031

QY 1067 GGCATGGTGACACGCTTTCACACAAATGCCAGGATGCCCTTAATGTCCTCCAGCT 1126
DB 1032 GGCATGGTGACACGCTTTCACACAAATGCCAGGATGCCCTTAATGTCCTCCAGCT 1091
QY 1127 TGGCGCTGGTCTTGGCCCTATGAAGAAGTTCTCGACTTCAGAAAGCTGGCGATGCC 1186
DB 1092 TGGCGCTGGTCTTGGCCCTATGAAGAAGTTCTCGACTTCAGAAAGCTGGCGATGCC 1151
QY 1187 ATGAATTTGCTGGGCTACCGCCAGTCAGATCTGAACAAGAACAGAAACCTGTGCTG 1246
DB 1152 ATGAATTTGCTGGGCTACCGCCAGTCAGATCTGAACAAGAACAGAAACCTGTGCTG 1211
QY 1247 GATCTTCTGTCTACCTGAGCTGTCCCTGAGCAATCCACTAAGAGGTTGAGAAGCTTT 1306
DB 1212 GATCTTCTGTCTACCTGAGCTGTCCCTGAGCAATCCACTAAGAGGTTGAGAAGCTTT 1271
Y 1307 GCGGCACCTGTGCTGACGCTCAGTCACTTCTCTGAATGCTTCTGAGCTTGCCTACAT 1366
DB 1272 GCGGCACCTGTGCTGACGCTCAGTCACTTCTCTGAATGCTTCTGAGCTTGCCTACAT 1331
QY 1367 CT 1368
DB 1332 CT 1333

RESULT 12
AR310485 2065 bp DNA linear PAT 12-JUN-2003
LOCUS AR310485 Sequence 26 from patent US 6558935.
DEFINITION AR310485
ACCESSION AR310485
VERSION AR310485.1 GI:31703448
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 2065)
Tang,Y.T., Corley,N.C., Guegler,K.J., Baughn,M.R., Lal,P., Yue,H.,
Hillman,J.L. and Azimzal,Y.
TITLE Human transferase protein
JOURNAL Patent: US 6558935-A 26 06-MAY-2003;
FEATURES Location/Qualifiers
source 1..2065
BASE COUNT 466 a 575 c 491 g 533 t
ORIGIN
Query Match 56.4%; Score 1153; DB 6; Length 2065;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1593; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 106 CAAGGCTTTCACACTTCAGCAATGCTACTGCTTAAAAAATGAGCTCCTGCTTCT 165
DB 152 CAAGGCTTTCACACTTCAGCAATGCTACTGCTTAAAAAATGAGCTCCTGCTTCT 211
QY 166 GGTTCCTCCAGATGGGCACTTGGCTTATTTCTTCCACATGTACAGCCACACATCAGCTC 225
DB 212 GGTTCCTCCAGATGGGCACTTGGCTTATTTCTTCCACATGTACAGCCACACATCAGCTC 271
QY 226 CCTGTATGAAGGACAGCCGAGGCGATGACAGTGCTGTCTGTCTCCGAGGCTC 285
DB 272 CCTGTATGAAGGACAGCCGAGGCGATGACAGTGCTGTCTGTCTCCGAGGCTC 331
QY 286 TGGCTCTCTTTTGGGCAAGCTTTTGGGCAAGCCAGATGTTTCTACCTGATGA 345
DB 332 TGGCTCTCTTTTGGGCAAGCTTTTGGGCAAGCCAGATGTTTCTACCTGATGA 391
QY 346 GCGCGCTGGCAGCTGTGATGACCTTCAAGCAGACACCGCTGATGCTGCATGGC 405
DB 392 GCGCGCTGGCAGCTGTGATGACCTTCAAGCAGACACCGCTGATGCTGCATGGC 451
QY 406 TGGCGGATGATGATGAGGCGCTTCTTGTGCGCATGAGGCTCTTGTGATGCTACAT 465

DB 452 TGTGCGGATCTGATACGGGCGCTCTTCTTGTGCGACATGAGGCTCTTGTGATGCTACAT 511
QY 466 GGAACCTGGTCCCGGAGACAGTCCAGCCCTTTCAGTGGGAGAAAGCGGGCCCTGTG 525
DB 512 GGAACCTGGTCCCGGAGACAGTCCAGCCCTTTCAGTGGGAGAAAGCGGGCCCTGTG 571
QY 526 TTCTGCACTGCTGTGACATCATCCACAAGATGAATATATCCCGGGCTCAGTCAG 585
DB 572 TTCTGCACTGCTGTGACATCATCCACAAGATGAATATATCCCGGGCTCAGTCAG 630
QY 586 GCTTCGTGTGAGTCAACAGCCCTTTG-AGTGGTGGAGAAAGGCTGCGCTCTACAGCC 644
DB 631 GCTTCGTGTGAGTCAACAGCCCTTTGAGTGGAGAAAGGCTGCGCTCTCTACAGCC 690
QY 645 AGCTGTGCTCAAGAGAGTGGCTTCTTCAACTGACAGTCCCTACCGGCTGCTGAAG 704
DB 691 AGCTGTGCTCAAGAGAGTGGCTTCTTCAACTGACAGTCCCTACCGGCTGCTGAAG 750
QY 705 ACCCTTCCTCAACCTGCATATGCTGACCTGTGCGGAGCCCGGGCCGTGTCGTT 764
DB 751 ACCCTTCCTCAACCTGCATATGCTGACCTGTGCGGAGCCCGGGCCGTGTCGTT 810
QY 765 CCGGAGACGCAAAAGGAGATCTCATGATTACAGTGGCATTTGTGATGGGACACATG 824
DB 811 CCGGAGACGCAAAAGGAGATCTCATGATTACAGTGGCATTTGTGATGGGACACATG 870
QY 825 AGCAAAATCTCAAGAGAGAGACCAACCTCATATGTGATGACAGTATGCTGCAAAAGCC 884
DB 871 AGCAAAATCTCAAGAGAGAGACCAACCTCATATGTGATGACAGTATGCTGCAAAAGCC 930
QY 885 AGCTGAGATCTCAAGAGACATCCAGTCTTGGCCCAAGGCGCTGACAGAAAGCTTACCTGC 944
DB 931 AGCTGAGATCTCAAGAGACATCCAGTCTTGGCCCAAGGCGCTGACAGAAAGCTTACCTGC 990
QY 945 TTGTGGGCTGTGAGGACCTGCTGAGCCCTGTGGGCCAGACTTCCGCAATGTATGAAT 1004
DB 991 TTGTGGGCTGTGAGGACCTGCTGAGCCCTGTGGGCCAGACTTCCGCAATGTATGAAT 1050
QY 1005 TCGTGGGATTTGAATTTCTGCCCCATCTTCAGACCTGTGGGTGATTAATCATCCCGAGGCA 1064
DB 1051 TCGTGGGATTTGAATTTCTGCCCCATCTTCAGACCTGTGGGTGATTAATCATCCCGAGGCA 1110
QY 1065 AGGCGATGGTGTACACAGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGG 1124
DB 1111 AGGCGATGGTGTACACAGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGG 1170
QY 1125 CTTGGCGCTGTGTTGGCCCTATGAAGGTTTCTGCACTTCAAGAAAGCTGTGGCGATG 1184
DB 1171 CTTGGCGCTGTGTTGGCCCTATGAAGGTTTCTGCACTTCAAGAAAGCTGTGGCGATG 1230
QY 1185 CCATGAATTTGCTGGGCTACCGCCAGCTCAGATCTGAACAAGAACAGAGAAACCTGTGCG 1244
DB 1231 CCATGAATTTGCTGGGCTACCGCCAGCTCAGATCTGAACAAGAACAGAGAAACCTGTGCG 1290
QY 1245 TGGATCTTCTGTACTTGGATGCTGCTGCTGAGCAATCCACTAAGAGGTTGAGAAAGCT 1304
DB 1291 TGGATCTTCTGTACTTGGATGCTGCTGCTGAGCAATCCACTAAGAGGTTGAGAAAGCT 1350
QY 1305 TTGCTGCACTGCTGTGACAGCTCAGTCACTTCTGCAATGCTTCTGAGACCTTGCCTAC 1364
DB 1351 TTGCTGCACTGCTGTGACAGCTCAGTCACTTCTGCAATGCTTCTGAGACCTTGCCTAC 1410
QY 1365 ATCTGTAGGCTTAACATACATGCTGTGGGTATCACAGTGAAGTGTGTCACAC 1424
DB 1411 ATCTGTAGGCTTAACATACATGCTGTGGGTATCACAGTGAAGTGTGTCACAC 1470
QY 1425 GTGCTCAAGCAGAGAGACTTTTGTGTCATGCTTGTGTCTAAGAAACAGACTGGGAAAGC 1484
DB 1471 GTGCTCAAGCAGAGAGACTTTTGTGTCATGCTTGTGTCTAAGAAACAGACTGGGAAAGC 1530
QY 1485 TTATGTAGCAGACATCCACAGAGTGAAGAAAGAGGATATGCTTCTTCTTCTTCTGATG 1544
DB 1531 TTATGTAGCAGACATCCACAGAGTGAAGAAAGAGGATATGCTTCTTCTTCTTCTGATG 1590

QY 1545 TTCTGTCTGGGAGACTTTCAGAGCTTTGTGGCTGGAGGCTTTAAGCAGCAGACAG 1604
 DB 1591 TTCCGTCTGGGAGACTTTCAGAGCTTTGTGGCTGGAGGCTTTAAGCAGCAGACAG 1650
 QY 1605 TATCAGTGAATGTATTCATTAACCTCCCTGTCCACATCTTGGCCCAATGGGAATGATC 1664
 DB 1651 TATCAGTGAATGTATTCATTAACCTCCCTGTCCACATCTTGGCCCAATGGGAATGATC 1710
 QY 1665 TTTCACCAAGAGCTTCACAGCATTTTCCACAGATCC 1703
 DB 1711 TTTCACCAAGAGCTTCACAGCATTTTCCACAGATCC 1749

RESULT 13

LOCUS BD124832 877 bp DNA linear PAT 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD124832
 VERSION BD124832.1 GI:23219777
 KEYWORDS JP 2002017375-A/263.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y., Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and Koga,H.
 TITLE Primer for synthesizing full-length cDNA and use thereof
 JOURNAL Patent: JP 2002017375-A 263 22-JAN-2002;
 HELIX RESEARCH INSTITUTE

COMMENT

PN OS Homo sapiens (human)
 PD JP 2002017375-A/263
 PF 22-JAN-2002
 PI 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA, TEISUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
 PI ISHII,
 PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TEISUTSU OTSUKI, HISASHI KOGA

PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/10,
 PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 Location/Qualifiers
 FT source
 1.877
 /organism="Homo sapiens (human)"
 /location="Qualifiers"
 1.877
 /organism="Homo sapiens (human)"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

FEATURES

BASE COUNT 183 a 266 c 221 g 203 t 4 others
 ORIGIN

Query Match 34.0%; Score 695; DB 6; Length 877;
 Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;
 Matches 695; Conservative 0;

QY 106 CAAGTCTTCCACTTCAGACACATGCTACTGCTAAATAATGAAGCTCTGCTTTCT 165
 DB 47 CAAGTCTTCCACTTCAGACACATGCTACTGCTAAATAATGAAGCTCTGCTTTCT 106
 QY 166 GGTTCCTCAGATGGCAGATCTTGGCTCTATTCTTCACATGTACAGCCCAACATCAGCTC 225
 DB 107 GGTTCCTCAGATGGCAGATCTTGGCTCTATTCTTCACATGTACAGCCCAACATCAGCTC 166
 QY 226 CCTGTCTATGAAGGACAGCCCGAGCGCATGACAGCTGCTGCTGCTTCTCTGCGGCTC 285
 DB 167 CCTGTCTATGAAGGACAGCCCGAGCGCATGACAGCTGCTGCTGCTTCTCTGCGGCTC 226

QY 286 TGGCTCTCTTTTGTGGGGAGCTTTTGGGAGCAGCCAGATGTTTTCTACCTGATGGA 345
 DB 227 TGGCTCTCTTTTGTGGGGAGCTTTTGGGAGCAGCCAGATGTTTTCTACCTGATGGA 286
 QY 346 GCCGCGCTGGCAGCTGTGGATGACCTTCAAGCAGACACCCTGGATGCTGACATGGC 405
 DB 287 GCCGCGCTGGCAGCTGTGGATGACCTTCAAGCAGACACCCTGGATGCTGACATGGC 346
 QY 406 TGTGGGATCTGATACAGGCGCCTTCTTGGGACATGAGCCCTTGTGATGCTACAT 465
 DB 347 TGTGGGATCTGATACAGGCGCCTTCTTGGGACATGAGCCCTTGTGATGCTACAT 406
 QY 466 GGAACCTGTCCTCCGAGACAGTCCAGGCTCTTTCAGTGGGAGAACACCGGCTCTG 525
 DB 407 GGAACCTGTCCTCCGAGACAGTCCAGGCTCTTTCAGTGGGAGAACACCGGCTCTG 466
 QY 526 TTCTGACCTGCTGTGACATCATCCACAGATGAATATCCCGGCTCTGCTGAG 585
 DB 467 TTCTGACCTGCTGTGACATCATCCACAGATGAATATCCCGGCTCTGCTGAG 526
 QY 586 GCTCTGTGACATCAACAGCCCTTTGAGTGGTGGAGAAAGCCCTGCTCTACAGCA 645
 DB 527 GCTCTGTGACATCAACAGCCCTTTGAGTGGTGGAGAAAGCCCTGCTCTACAGCA 586
 QY 646 GCTGTGCTCAAGAGAGTGGCGCTTCTCAACCTGAGTCCCTCAACCTGCTGAAGA 705
 DB 587 GCTGTGCTCAAGAGAGTGGCGCTTCTCAACCTGAGTCCCTCAACCTGCTGAAGA 646
 QY 706 CCCCTGCTCAACCTGATATGCTGACCTGCTGGGAGCCCGGCGCTGCTGCTG 765
 DB 647 CCCCTGCTCAACCTGATATGCTGACCTGCTGGGAGCCCGGCGCTGCTGCTG 706
 QY 766 CCGAGAACGCAAGAGGAGATCTATATTGACA 800
 DB 707 CCGAGAACGCAAGAGGAGATCTATATTGACA 741

RESULT 14

LOCUS BD126950 877 bp DNA linear PAT 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD126950
 VERSION BD126950.1 GI:23221895
 KEYWORDS JP 2002017375-A/2381.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y., Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and Koga,H.
 TITLE Primer for synthesizing full-length cDNA and use thereof
 JOURNAL Patent: JP 2002017375-A 2381 22-JAN-2002;
 HELIX RESEARCH INSTITUTE

COMMENT

PN OS Homo sapiens (human)
 PD JP 2002017375-A/2381
 PF 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA, TEISUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
 PI ISHII,
 PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TEISUTSU OTSUKI, HISASHI KOGA

PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/10,
 PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 Location/Qualifiers
 FT source
 1.877
 /organism="Homo sapiens (human)"
 /location="Qualifiers"

FEATURES

```

source 1. .877
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
BASE COUNT 183 a 266 c 221 g 203 t 4 others
ORIGIN
Query Match 34.0%; Score 695; DB 6; Length 877;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 CAAGCTCTCCATTGACGACATGCTACTGCTAAATAAAGTCTCTGCTTTCT 165
Db 47 CAAGCTCTCCATTGACGACATGCTACTGCTAAATAAAGTCTCTGCTTTCT 106
QY 166 GGTTCCTCCAGATGGCCATCTGGCTCTATCTCTCCACATGTACAGCCACAATCAGCTC 225
Db 107 GGTTCCTCCAGATGGCCATCTGGCTCTATCTCTCCACATGTACAGCCACAATCAGCTC 166
QY 226 CCTGTCTATGAAAGGACACAGCCGAGGACATGACAGTGTCTGTCTCTCTGGGCTC 285
Db 167 CCTGTCTATGAAAGGACACAGCCGAGGACATGACAGTGTCTGTCTCTCTGGGCTC 226
QY 286 TGCTCTCTCTTTGTGGGAGCTTTTGGGAGACCCAGATGTTTCTACCTGATGA 345
Db 227 TGCTCTCTCTTTGTGGGAGCTTTTGGGAGACCCAGATGTTTCTACCTGATGA 286
QY 346 GCCCGCTGGACAGTGTGATGACCTTCAAGACAGACCGCTGATGCTGCATGGC 405
Db 287 GCCCGCTGGACAGTGTGATGACCTTCAAGACAGACCGCTGATGCTGCATGGC 346
QY 406 TGTGGGGATCTGATACGGGCGCTCTTGTGCGACATGAGGCTCTTATGCTATCAT 465
Db 347 TGTGGGGATCTGATACGGGCGCTCTTGTGCGACATGAGGCTCTTATGCTATCAT 406
QY 466 GGAACCTGTGTCCCGGAGACAGTCCAGCCCTTTCAGTGGGAGAAAGCCGGCCCTGTG 525
Db 407 GGAACCTGTGTCCCGGAGACAGTCCAGCCCTTTCAGTGGGAGAAAGCCGGCCCTGTG 466
QY 526 TTGTGCACTGCTGTGACATCATCCACAAGATGAATCATCCCGGGGCTCAGTCAG 585
Db 467 TTGTGCACTGCTGTGACATCATCCACAAGATGAATCATCCCGGGGCTCAGTCAG 526
QY 586 GCTCCTGTGAGTCAACAGCCCTTGTAGGTGTGGGAAAGGCGCTGCCCTCCACAGCA 645
Db 527 GCTCCTGTGAGTCAACAGCCCTTGTAGGTGTGGGAAAGGCGCTGCCCTCCACAGCA 586
QY 646 CGTGTGTCTCAAGAGAGTGGCTTCTTCAACTGCAAGTCCCTTACCCGCTGCTGAAGA 705
Db 587 CGTGTGTCTCAAGAGAGTGGCTTCTTCAACTGCAAGTCCCTTACCCGCTGCTGAAGA 646
QY 706 CCCCTCCCTCAACCTGATATCGTGCACCTGATCGGGACCCCGGGCGCTGTTCCGTTT 765
Db 647 CCCCTCCCTCAACCTGATATCGTGCACCTGATCGGGACCCCGGGCGCTGTTCCGTTT 706
QY 766 CCGAGAACGACAAAGGAGATGTCATGATTGACA 800
Db 707 CCGAGAACGACAAAGGAGATGTCATGATTGACA 741

RESULT 15
AX381256 517 bp DNA linear PAT 18-MAR-2002
LOCUS AX381256
DEFINITION Sequence 194 from Patent W00212280.
ACCESSION AX381256
VERSION AX381256.1 GI:19576075
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Pyle,R.A., Xu,J. and Secretist,H.

```

```

TITLE Compositions and methods for the therapy and diagnosis of colon
        cancer
JOURNAL Patent: WO 0212280-A 194 14-FEB-2002;
CORIXA CORPORATION (US)
FEATURES
        source 1. .517
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
BASE COUNT 121 a 130 c 124 g 142 t
ORIGIN
Query Match 19.7%; Score 403; DB 6; Length 517;
Best Local Similarity 100.0%; Pred. No. 3.5e-228;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GGCTTGGCTGCGACCGTGGTGCAGCTCAGTCACTTCTCTGTAATGCTTGTAGCCCTGC 1360
Db 13 GGCTTGGCTGCGACCGTGGTGCAGCTCAGTCACTTCTCTGTAATGCTTGTAGCCCTGC 72
QY 1361 CTACATCTGAGACCTTAACATGATGCTGTGGTATACACATGAGTGTGATGTCC 1420
Db 73 CTACATCTGAGACCTTAACATGATGCTGTGGTATACACATGAGTGTGATGTCC 132
QY 1421 ACACGTGCTAAGACAGAGACTTGTGTGCTCATGCTGTGTCTAGAAAAACAGCTGGG 1480
Db 133 ACACGTGCTAAGACAGAGACTTGTGTGCTCATGCTGTGTCTAGAAAAACAGCTGGG 192
QY 1481 AACCTTATGAGACAGACATCCACACAGAGAAAGGATGCTCTTCTTCTTCTT 1540
Db 193 AACCTTATGAGACAGACATCCACACAGAGAAAGGATGCTCTTCTTCTTCTTCTT 252
QY 1541 GATCTTCTGTGTGGGACAGCTTCAAGAGACTTGTGGCTGAGGCCATTAAAGACAGAC 1600
Db 253 GATCTTCTGTGTGGGACAGCTTCAAGAGACTTGTGGCTGAGGCCATTAAAGACAGAC 312
QY 1601 ACAGTATCACTGGAATGATGATCCATTAACCTCCCTGTCCACATCTTGGCCCAATGGGAATG 1660
Db 313 ACAGTATCACTGGAATGATGATCCATTAACCTCCCTGTCCACATCTTGGCCCAATGGGAATG 372
QY 1661 GATCTTCAACAAAGAGCTCACCAGATTTTCCACAGAGATGC 1703
Db 373 GATCTTCAACAAAGAGCTCACCAGATTTTCCACAGAGATGC 415

Search completed: August 14, 2003, 23:39:20
Job time : 7286 secs

```

[illegible]


```

Db      |||||
340 TGGCTCTCTTTTGGGGGACCTTTTGGGAGACCCAGATGTTTCTACCTGATGGA 399
QY      GCGCCGCTGGGACAGTGTGATGACCTTCAAGCAGAGACACGGCTGGATGCTGACATGGC 405
Db      GCGCCGCTGGGACAGTGTGATGACCTTCAAGCAGAGACACGGCTGGATGCTGACATGGC 459
QY      TGTGGGGGATGTATACGGGGCGTCTTGTGCGACATGAGCGTCTTGTATGCTTACAT 465
Db      TGTGGGGGATGTATACGGGGCGTCTTGTGCGACATGAGCGTCTTGTATGCTTACAT 519
QY      GGAACCTGGTCCCGGAGACATGCAAGCTTTTCAAGTGGGAGAACAGCCGGGCGCTGTG 525
Db      GGAACCTGGTCCCGGAGACATGCAAGCTTTTCAAGTGGGAGAACAGCCGGGCGCTGTG 579
QY      TTCTGACCTGGCTGTGATCATGCTCCACAAAGATGAATCATCCCGGGGCTCAGTCAG 585
Db      TTCTGACCTGGCTGTGATCATGCTCCACAAAGATGAATCATCCCGGGGCTCAGTCAG 639
QY      GCTCTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 645
Db      GCTCTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 699
QY      CGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 705
Db      CGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 759
QY      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 819
Db      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 825
QY      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 879
Db      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 885
QY      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 939
Db      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 945
QY      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 999
Db      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 1005
QY      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 1059
Db      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 1065
QY      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 1119
Db      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 1125
QY      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 1179
Db      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 1185
QY      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 1239
Db      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 1245
QY      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 1299
Db      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 1305
QY      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 1359
Db      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 1365
QY      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 1419
Db      CCGTGTGTGTGACATCAACAGCCCTTGTAGAGTGTGAGAGAGCCCTGCGCTCTACAGCCA 1425

```

```

Db      1420 TCTTGAGGCTTAACTACATGCTGTGGGTATTCACACTGAGTGTGATGTTGTCACAGC 1479
QY      1426 TGTCTAAGCAGAGAGACTTTTGTGTCATGCTTGTGTAGAAAAAGAGATGGGAACTT 1485
Db      1480 TGTCTAAGCAGAGAGACTTTTGTGTCATGCTTGTGTAGAAAAAGAGATGGGAACTT 1539
QY      1486 TATGTAGAGCAGACATCCACAGTAGAGTAAGAGGATTTGCTCTTCTTCTTCTTGTATCT 1545
Db      1540 TATGTAGAGCAGACATCCACAGTAGAGTAAGAGGATTTGCTCTTCTTCTTCTTGTATCT 1599
QY      1546 TCTCTGTGTGAGCAGATTTAGAGACTTTTGGCTGTGAGAGCCCTATTAACAGACAGACT 1605
Db      1600 TCTCTGTGTGAGCAGATTTAGAGACTTTTGGCTGTGAGAGCCCTATTAACAGACAGACT 1659
QY      1606 ATCAGTGAATGATTCATTAACCTCCCTGTCCATCTTGGCCAAATGGGAAATGATCT 1665
Db      1660 ATCAGTGAATGATTCATTAACCTCCCTGTCCATCTTGGCCAAATGGGAAATGATCT 1719
QY      1666 TTCCACCAAGAGCTCACAGACTTTTTCACAGAGATGC 1703
Db      1720 TTCCACCAAGAGCTCACAGACTTTTTCACAGAGATGC 1757

RESULT 2
US-10-007-262-2
; Sequence 2, Application US/10007262
; Publication No. US20020164748A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/10/007, 262
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190, 911
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens
US-10-007-262-2

Query Match      75.7%; Score 1547; DB 13; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      106 CAAGGCTTTCACATTCAGCACAATGCTACTGCTTAAATAAATGAAGCTCTGCTGTTCT 165
Db      160 CAAGGCTTTCACATTCAGCACAATGCTACTGCTTAAATAAATGAAGCTCTGCTGTTCT 219
QY      166 GGTTCCTCCAGATGGCCATTTGGCTTATTTTCCACATGTACAGCCACAATATCAGCTC 225
Db      220 GGTTCCTCCAGATGGCCATTTGGCTTATTTTCCACATGTACAGCCACAATATCAGCTC 279
QY      226 CCGTGTATGAAGCAGACCCGAGGAGCATGTGACAGGCTGCTTCTTCTGAGGCTC 285
Db      280 CCGTGTATGAAGCAGACCCGAGGAGCATGTGACAGGCTGCTTCTTCTGAGGCTC 339
QY      286 TGGCTCTTCTTTTGTGGGAGACTTTTGGGAGCACCAGATGTTTCTACCTGATGGA 345
Db      340 TGGCTCTTCTTTTGTGGGAGACTTTTGGGAGCACCAGATGTTTCTACCTGATGGA 399
QY      346 GCGCCGCTGGGACAGTGTGATGACCTTCAAGCAGAGACACCGCTGATGCTGACATGGC 405
Db      400 GCGCCGCTGGGACAGTGTGATGACCTTCAAGCAGAGACACCGCTGATGCTGACATGGC 459
QY      406 TGTGGGGATGTGATACGGGGCGTCTTGTGCGACATGAGAGCGTCTTGTATGCTTACAT 465
Db      460 TGTGGGGATGTGATACGGGGCGTCTTGTGCGACATGAGAGCGTCTTGTATGCTTACAT 519

```


QY 466 GGAACCTGTCCTCCGAGAGAGTCACGCTCTTTCAGTGGGAGAAACAGCCGGCCCTGTG 525
 Db 520 GGAACCTGTCCTCCGAGAGAGTCACGCTCTTTCAGTGGGAGAAACAGCCGGCCCTGTG 579
 QY 526 TTCTGACACCTGCTGTGATCATCCACAAGATGAATCAATCCCGGGCTCACTGAG 585
 Db 580 TTCTGACACCTGCTGTGATCATCCACAAGATGAATCAATCCCGGGCTCACTGAG 639
 QY 586 GCTCCCTGACGTCACAGCCCTTGTAGGTGTGTGAGAGAGCCCTGCGCTCTACAGCA 645
 Db 640 GCTCCCTGACGTCACAGCCCTTGTAGGTGTGTGAGAGAGCCCTGCGCTCTACAGCA 699
 QY 646 CGTGTGCTCAAGAGAGTGGCTCTTCAACCTGACGTCCCTTACCCGCTGTGAAGA 705
 Db 700 CGTGTGCTCAAGAGAGTGGCTCTTCAACCTGACGTCCCTTACCCGCTGTGAAGA 759
 QY 706 CCCCTCCCTCAACCTGATATCGTACCTGTGTCCGGAGCCCGGGCCGTTCGCTTC 765
 Db 760 CCCCTCCCTCAACCTGATATCGTACCTGTGTCCGGAGCCCGGGCCGTTCGCTTC 819
 QY 766 CCGAGAGGACAAAGGAGATCATGATGAGTTCGATGAGTGGGAGCATGA 825
 b 820 CCGAGAGGACAAAGGAGATCATGATGAGTTCGATGAGTGGGAGCATGA 879
 QY 826 GCAAAAACCTCAAGAGAGAGACCAACCTTACTATGTATGATGAGTTCGCAAAAGCA 885
 Db 880 GCAAAAACCTCAAGAGAGAGACCAACCTTACTATGTATGATGAGTTCGCAAAAGCA 939
 QY 886 GCTGTGATCTACAAAGCATCCAGTCTTGTCCCAAGCCCTGACAGAGAGCATGCT 945
 Db 940 GCTGTGATCTACAAAGCATCCAGTCTTGTCCCAAGCCCTGACAGAGAGCATGCT 999
 QY 946 TGTGCGCTATGAGAGACCTGGCTGAGCCCTGTGTGCCAGACCTCCGGAATGTGAT 1005
 Db 1000 TGTGCGCTATGAGAGACCTGGCTGAGCCCTGTGTGCCAGACCTCCGGAATGTGAT 1059
 QY 1006 CGTGTGATGGAATCTTGTGCCCATCTTTCAGACCTGGGTGATTAATCAACCCAGAGCA 1065
 Db 1060 CGTGTGATGGAATCTTGTGCCCATCTTTCAGACCTGGGTGATTAATCAACCCAGAGCA 1119
 QY 1066 GGGCATGGTGACCAACCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGC 1125
 Db 1120 GGGCATGGTGACCAACCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGC 1179
 QY 1126 TTGGCGCTGCTTGTGCCATATGAAAGTTCCTGCACTTCAGAAACCTGTGGCATGC 1185
 Db 1180 TTGGCGCTGCTTGTGCCATATGAAAGTTCCTGCACTTCAGAAACCTGTGGCATGC 1239
 QY 1186 CATGAATTTGCTGGGCTACCGCCACGTCAATCTGAACAAGAAACAGAAACCTGTGCT 1245
 b 1240 CATGAATTTGCTGGGCTACCGCCACGTCAATCTGAACAAGAAACAGAAACCTGTGCT 1299
 QY 1246 GGAATCTTCTGTACCTGAGCTGTCCCTGAGCAAAATCCACTAAGAGGGTGAAGAGCT 1305
 Db 1300 GGAATCTTCTGTCTACTGTGAGCTGTCCCTGAGCAAAATCCACTAAGAGGGTGAAGAGCT 1359
 QY 1306 TGTGCGACCTGCTGTGAGCTGTGACCTTCTCTGAATGCTTCTGAGCTTCCCTACA 1365
 Db 1360 TGTGCGACCTGCTGTGAGCTGTGACCTTCTCTGAATGCTTCTGAGCTTCCCTACA 1419
 QY 1366 TCTCTGAGCTTAATCATGATGTCTGTGGTATCATCACTGATGAGTGTGTGTCACAG 1425
 Db 1420 TCTCTGAGCTTAATCATGATGTCTGTGGTATCATCACTGATGAGTGTGTGTCACAG 1479
 QY 1426 TGTCTCAAGCAGAAAGACTTTTGTGTCCATGCTGTGTCTAGAAAACAGAGTGGGAACT 1485
 Db 1480 TGTCTCAAGCAGAAAGACTTTTGTGTCCATGCTGTGTCTAGAAAACAGAGTGGGAACT 1539
 QY 1486 TATGTGAGCAGACATCCCAAGGATGAAGAGGATATGCTCTTCTTCTTGTGATCT 1545
 Db 1540 TATGTGAGCAGACATCCCAAGGATGAAGAGGATATGCTCTTCTTCTTGTGATCT 1599

QY 1546 TCTGTCTGGGACAGACTTCAGAGACTTTGTGGCTGTGAGGCTTATTAAGCAGACAGT 1605
 Db 1600 TCTGTCTGGGACAGACTTCAGAGACTTTGTGGCTGTGAGGCTTATTAAGCAGACAGT 1659
 QY 1606 ATCAGTGAATTTGATTCATTAACCTCCCTGTCCACATCTTCCCAATGGGAATGATCT 1665
 Db 1660 ATCAGTGAATTTGATTCATTAACCTCCCTGTCCACATCTTCCCAATGGGAATGATCT 1719
 QY 1666 TTCACAAAGAGCTCACCAGATTTTCCACAGAGATGC 1703
 Db 1720 TTCACAAAGAGCTCACCAGATTTTCCACAGAGATGC 1757

RESULT 3

US-09-598-2595
 ; Sequence 2595, Application US/0998598
 ; Patent No. US20020150922A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Xu, Jlangchun
 ; APPLICANT: Chenault, Ruth A.
 ; APPLICANT: Mesinger, Madelein Joy
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE REFERENCE: 210121.561
 ; CURRENT FILING DATE: 2001-11-16
 ; NUMBER OF SEQ ID NOS: 2606
 ; SOFTWARE: Cortix Invention Disclosure Database
 ; SEQ ID NO 2595
 ; LENGTH: 505
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-598-2595

Query Match 19.7%; Score 403; DB 10; Length 505;
 Best Local Similarity 100.0%; Pred. No. 8.5e-203;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GGCCTTGTGCTGACCTGCTGTCACAGCTTCACATCTTCTGTGAATGCTTGTGAGCCCTTGC 1360
 Db 1 GGCCTTGTGCTGACCTGCTGTCACAGCTTCACATCTTCTGTGAATGCTTGTGAGCCCTTGC 60
 QY 1361 CTACATCTCTAGGCTTAACATACATGCTGTGGGTATCACAGTGTGATGCTGTGTC 1420
 Db 61 CTACATCTCTAGGCTTAACATACATGCTGTGGGTATCACAGTGTGATGCTGTGTC 120
 QY 1421 ACAGTGTCTCAAGCAGAAAGACTTTTGTGTCCATGCTGTGTCTAGAAAACAGAGCTGGG 1480
 Db 121 ACAGTGTCTCAAGCAGAAAGACTTTTGTGTCCATGCTGTGTCTAGAAAACAGAGCTGGG 180
 QY 1481 AACCTTATGTGAGCAGACATCCACAGTGAAGAGGGTATGCTCTTCTTCTTCTTCTT 1540
 Db 181 AACCTTATGTGAGCAGACATCCACAGTGAAGAGGGTATGCTCTTCTTCTTCTTCTT 240
 QY 1541 GATCTTCCTGTCTGGGAGACTTCAGAGACTTTGTGGCTGTGAGGCTTATTAAGCAGAGC 1600
 Db 241 GATCTTCCTGTCTGGGAGACTTCAGAGACTTTGTGGCTGTGAGGCTTATTAAGCAGAGC 300
 QY 1601 ACAGTATCAGTGAATTTGATTCATTAACCTCCCTGTCCACATCTTGTGCCAATGGGGAATG 1660
 Db 301 ACAGTATCAGTGAATTTGATTCATTAACCTCCCTGTCCACATCTTGTGCCAATGGGGAATG 360
 QY 1661 GATCTTTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
 Db 361 GATCTTTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 403

RESULT 4

US-09-580-194
 ; Sequence 194, Application US/09919580
 ; Patent No. US20020110832A1
 ; GENERAL INFORMATION:

APPLICANT: Pyle, Ruth
APPLICANT: Xu, Jiangchun
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.552
CURRENT APPLICATION NUMBER: US/09/919,580
CURRENT FILING DATE: 2001-07-30
NUMBER OF SEQ ID NOS: 934
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 194
LENGTH: 517
TYPE: DNA
ORGANISM: Homo sapiens
US-09-919-580-194

Query Match 19.7% Score 403 DB 10 Length 517
Best Local Similarity 100.0% Pred. No. 8.5e-203 Indels 0 Gaps 0
Matches 403 Conservative 0 Mismatches 0

QY 1301 GGCTTTGCTGCGACCTGCTGAGCCTCAGCTCAGCTTTCTGTGAAATGCTTGTGAGCCTTGC 1360
DB 13 GGCTTTGCTGCGACCTGCTGAGCCTCAGCTCAGCTTTCTGTGAAATGCTTGTGAGCCTTGC 72
QY 1361 CTACATCTCTGAGCCTTAATCTACATCTCTGTGGGTATCAGACTGAGTGTGTCTCC 1420
DB 73 CTACATCTCTGAGCCTTAATCTACATCTCTGTGGGTATCAGACTGAGTGTGTCTCC 132
QY 1421 ACAGCTGCTCAAGCAGAAAGAGCTTTGTCTCATGCTTGTGTCTAGAAAAGAGCTGGGG 1480
DB 133 ACAGCTGCTCAAGCAGAAAGAGCTTTGTGTCTCATGCTTGTGTCTAGAAAAGAGCTGGGG 192
QY 1481 AACCTTATGTGAGCAGCAGCATCCACAGCTGAAAGAGGCTATTTCTCTCTCTTTCTT 1540
DB 193 AACCTTATGTGAGCAGCAGCATCCACAGCTGAAAGAGGCTATTTCTCTCTCTTTCTT 252
QY 1541 GATCTTCCCTGCTGGGAGACATCTGAGAGCTTGTGCTGAGAGGCTTATTAAGCAGAC 1600
DB 253 GATCTTCCCTGCTGGGAGACATCTGAGAGCTTGTGCTGAGAGGCTTATTAAGCAGAC 312
QY 1601 ACAGTATCAGTGAATTCATTAACCTCCCTGTCCACATCTTCCCAATGGGGAGATG 1660
DB 313 ACAGTATCAGTGAATTCATTAACCTCCCTGTCCACATCTTCCCAATGGGGAGATG 372
QY 1661 GATCTTCCCTGCTGGGAGACATCTGAGAGCTTGTGCTGAGAGGCTTATTAAGCAGATGC 1703
DB 373 GATCTTCCCTGCTGGGAGACATCTGAGAGCTTGTGCTGAGAGGCTTATTAAGCAGATGC 415

RESULT 5
US-09-919-580-264
Sequence 264, Application US/09919580
Patent No. US20020110832A1
GENERAL INFORMATION:
APPLICANT: Pyle, Ruth
APPLICANT: Xu, Jiangchun
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.552
CURRENT APPLICATION NUMBER: US/09/919,580
CURRENT FILING DATE: 2001-07-30
NUMBER OF SEQ ID NOS: 934
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 264
LENGTH: 389
TYPE: DNA
ORGANISM: Homo sapiens
US-09-919-580-264

Query Match 5.6% Score 114 DB 10 Length 389
Best Local Similarity 100.0% Pred. No. 1.3e-49 Indels 0 Gaps 0
Matches 114 Conservative 0 Mismatches 0

QY 1930 TGCCATCTGTAACTAATAAATCCCAAAATGCTGTGTAATGTCCTTTTATG 1989
DB 266 TGCCATCTGTAACTAATAAATCCCAAAATGCTGTGTAATGTCCTTTTATG 325
QY 1990 CTCTTAATTAATAGCAGTAATGTCATTTTATGGATCTTAAAAA 2043
DB 326 CTCTTAATTAATAGCAGTAATGTCATTTTATGGATCTTAAAAA 379

RESULT 6
US-09-927-602-1
Sequence 1, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPLICANT: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2544
TYPE: DNA
ORGANISM: Homo Sapien
NAME/KEY: CDS
LOCATION: (693)...(1877)
US-09-927-602-1

Query Match 2.9% Score 59 DB 9 Length 2544
Best Local Similarity 100.0% Pred. No. 1.9e-20 Indels 0 Gaps 0
Matches 59 Conservative 0 Mismatches 0

QY 624 AGGCTGCGCGCTCTCTACAGCAGCTGCTCAAGAGGTGCGCTTCTCAACCTGCAG 682
DB 1180 AGGCTGCGCGCTCTCTACAGCAGCTGCTCAAGAGGTGCGCTTCTCAACCTGCAG 1238

RESULT 7
US-09-927-602-38
Sequence 38, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPLICANT: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 38
LENGTH: 48436
TYPE: DNA
ORGANISM: Homo Sapien
US-09-927-602-38

Query Match 2.9% Score 59 DB 9 Length 48436
Best Local Similarity 100.0% Pred. No. 1.9e-20 Indels 0 Gaps 0
Matches 59 Conservative 0 Mismatches 0

QY 624 AGGCTGCGCGCTCTCTACAGCAGCTGCTCAAGAGGTGCGCTTCTCAACCTGCAG 682
DB 48024 AGGCTGCGCGCTCTCTACAGCAGCTGCTCAAGAGGTGCGCTTCTCAACCTGCAG 48082

RESULT 8
US-09-816-825-7
; Sequence 7, Application US/09816825
; Patent No. US20010051370A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CON
; CURRENT APPLICATION NUMBER: US/09/816, 825
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/045, 284
; PRIOR FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-816-825-7

Query Match
Best Local Similarity 1.8%; Score 37; DB 9; Length 37;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 830 AACTCAAGAGGAGGACCAACCTACTATGTATGTC 866
Db 1 AACTCAAGAGGAGGAGGACCAACCTACTATGTATGTC 37

RESULT 9
US-10-007-262-7
; Sequence 7, Application US/10007262
; Publication No. US20020164748A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/10/007, 262
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190, 911
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: DNA
; ORGANISM: H. sapiens
US-10-007-262-7

Query Match
Best Local Similarity 1.8%; Score 37; DB 13; Length 37;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 830 AACTCAAGAGGAGGACCAACCTACTATGTATGTC 866
Db 1 AACTCAAGAGGAGGAGGACCAACCTACTATGTATGTC 37

RESULT 10
US-09-867-701-9458
; Sequence 9458, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
US-09-867-701-9458

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867, 701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9458
; LENGTH: 167
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9458

Query Match
Best Local Similarity 1.1%; Score 22; DB 10; Length 167;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAAGACGACAGAGGAGGTAGAG 44
Db 3 GAAGACGACAGAGGAGGTAGAG 24

RESULT 11
US-09-867-701-9067
; Sequence 9067, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867, 701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9067
; LENGTH: 171
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9067

Query Match
Best Local Similarity 1.1%; Score 22; DB 10; Length 171;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAAGACGACAGAGGAGGTAGAG 44
Db 9 GAAGACGACAGAGGAGGTAGAG 30

RESULT 12
US-09-867-701-9185
; Sequence 9185, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867, 701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9185
; LENGTH: 187
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9185

Query Match 1.1%; Score 22; DB 10; Length 187;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAAGACGACAGAGGCTAGAGG 44
DB 32 GAAGACGACAGAGGCTAGAGG 53

RESULT 13
US-09-867-701-9284
; Sequence 9284, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Agilate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9284
; LENGTH: 191
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9284

Query Match 1.1%; Score 22; DB 10; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAAGACGACAGAGGCTAGAGG 44
DB 29 GAAGACGACAGAGGCTAGAGG 50

RESULT 14
US-09-960-352-11099/C
; Sequence 11099, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengding
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11099
; LENGTH: 398
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 48-BOVMS1-008-Q1-E1-D8
US-09-960-352-11099

Query Match 1.0%; Score 20; DB 10; Length 398;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 GATGCCATCTGGCTCTAT 194
DB 240 GATGCCATCTGGCTCTAT 221

RESULT 15
US-10-027-632-140660
; Sequence 140660, Application US/10027632

; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 140660
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-140660

Query Match 1.0%; Score 20; DB 13; Length 570;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 GAAAGACCCCTCCCTCAACC 719
DB 273 GAAAGACCCCTCCCTCAACC 292

Search completed: August 14, 2003, 21:28:26
Job time: 539 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OK nucleic - nucleic search, using sw model

Run on: August 14, 2003, 17:44:36 ; Search time 144 Seconds
(without alignments)
6262.123 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043

Sequence: 1 gaattcattgtgttggtta.....tgggattcctaataaaaaaa 2043

Scoring table:

Gapop 60.0, Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1131652

Minimum DB seq length: 10

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

Issued_Patents_NA.*
1: /cgn2-6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2-6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2-6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2-6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2-6/ptodata/1/ina/PTCUTS.COMB.seq:*
6: /cgn2-6/ptodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1547	75.7	2032	3	US-09-045-284A-1
2	1547	75.7	2032	4	US-09-190-911-2
3	1153	56.4	2065	4	US-09-786-240-26
4	37	1.8	37	3	US-09-045-284A-7
5	37	1.8	37	4	US-09-190-911-7
6	19	0.9	3792	2	US-08-992-752-1
7	19	0.9	3792	2	US-08-302-752-1
8	19	0.9	5234	2	US-08-992-334-2
9	19	0.9	5234	2	US-08-302-752-2
10	19	0.9	6722	2	US-08-992-334-3
11	19	0.9	6722	2	US-08-302-752-3
12	18	0.9	223	2	US-08-778-494B-112
13	18	0.9	47	3	US-09-045-284A-8
14	18	0.9	47	4	US-09-190-911-8
15	18	0.9	150	4	US-09-439-313-320
16	18	0.9	150	4	US-09-352-616A-320
17	18	0.9	150	4	US-09-232-149A-320
18	18	0.9	201	3	US-09-020-956-78
19	18	0.9	201	3	US-09-030-607-78
20	18	0.9	201	4	US-09-439-313-78
21	18	0.9	201	4	US-09-352-616A-78
22	18	0.9	201	4	US-09-232-149A-78
23	18	0.9	202	3	US-09-030-607-180
24	18	0.9	202	4	US-09-439-313-180
25	18	0.9	202	4	US-09-352-616A-180
26	18	0.9	202	4	US-09-232-149A-180
27	18	0.9	536	3	US-09-020-956-69

28	18	0.9	536	3	US-09-030-607-69	Sequence 69, App1
29	18	0.9	536	4	US-09-439-313-69	Sequence 69, App1
30	18	0.9	536	4	US-09-352-616A-69	Sequence 69, App1
31	18	0.9	536	4	US-09-232-149A-69	Sequence 69, App1
32	18	0.9	822	4	US-09-107-532A-1333	Sequence 1333, App1
33	18	0.9	822	3	US-09-020-956-29	Sequence 29, App1
34	18	0.9	822	3	US-09-030-607-29	Sequence 29, App1
35	18	0.9	822	4	US-09-439-313-29	Sequence 29, App1
36	18	0.9	822	4	US-09-352-616A-29	Sequence 29, App1
37	18	0.9	822	4	US-09-232-149A-29	Sequence 29, App1
38	18	0.9	1197	4	US-09-257-179-26	Sequence 26, App1
39	18	0.9	1368	4	US-09-118-554-61	Sequence 61, App1
40	18	0.9	1368	4	US-09-118-627-61	Sequence 61, App1
41	18	0.9	1368	4	US-09-602-877A-61	Sequence 61, App1
42	18	0.9	1913	4	US-09-599-360B-40	Sequence 40, App1
43	18	0.9	1927	4	US-09-336-536-66	Sequence 66, App1
44	18	0.9	4351	4	US-09-634-238-25	Sequence 25, App1
45	17	0.8	141	1	US-08-036-555B-161	Sequence 161, App1

ALIGNMENTS

```
RESULT 1
US-09-045-284A-1
: Sequence 1, Application US/09045284A
: Patent No. 6265192
: GENERAL INFORMATION:
: APPLICANT: Bistrup, Annette
: APPLICANT: Rosen, Steven D.
: TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
: FILE REFERENCE: 6510-107US1
: CURRENT APPLICATION NUMBER: US/09/045,284A
: CURRENT FILING DATE: 1998-03-20
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 2032
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-045-284A-1

Query Match
Best Local Similarity 75.7%; Score 1547; DB 3; Length 2032;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 106 CAAGGTCCTCCACCTTCAGCACAATGCTACCTGCTTAAAAAATGAAGCTCTGCTGTTCT 165
    |||||
DB 160 CAAGGTCCTCCACCTTCAGCACAATGCTACCTGCTTAAAAAATGAAGCTCTGCTGTTCT 219

QY 166 GGTTCCTCCAGATGGGCACTTGGGCTCTATCTTCCACATGTCACCCACCAATCAGCTC 225
    |||||
DB 220 GGTTCCTCCAGATGGGCACTTGGGCTCTATCTTCCACATGTCACCCACCAATCAGCTC 279

QY 226 CCTGCTATGAAGCAGACCCGAGCAGATGCTGCTGTTGTTCTGTTCTGCGCTC 285
    |||||
DB 280 CCTGCTATGAAGCAGACCCGAGCAGATGCTGCTGTTGTTCTGTTCTGCGCTC 339

QY 286 TGGCTCTTCTTTTGGGGCAGCTTTTGGGCGACCCAGCCAGATGTTTCTACCTGATGGA 345
    |||||
DB 340 TGGCTCTTCTTTTGGGGCAGCTTTTGGGCGACCCAGCCAGATGTTTCTACCTGATGGA 399

QY 346 GCCCGGCGGACGCTGTCGATGATCACTTAAAGAGAGCAACCGCTGGATGTCGACATGGC 405
    |||||
DB 400 GCCCGGCGGACGCTGTCGATGATCACTTAAAGAGAGCAACCGCTGGATGTCGACATGGC 459

QY 406 TGTGCGGATCTGATACGGGCGCTTCTTGTGACATGAGCGCTTGTGATGCTACAT 465
    |||||
DB 460 TGTGCGGATCTGATACGGGCGCTTCTTGTGACATGAGCGCTTGTGATGCTACAT 519

QY 466 GGAACCTGTCCTCCGAGACAGTCAGCTCTTTTCAAGTGGGAGAAACAGCGGCGCTGTG 525
    |||||
```

D 520 GGAACCTGGTCCCGGAGACATCCAGCCCTTTCAGTGGGAAACACCGGGCCCTGTG 579
Q 526 TTCTGACACTGCTGTGACATCATCCACAAGATGAATATCCCGGGCTCAGTGCAG 585
D 580 TTCTGACACTGCTGTGACATCATCCACAAGATGAATATCCCGGGCTCAGTGCAG 639
Q 586 GCTCCGTGTGAGCAACAGCCCTTTGAGTGTGAGAAAGCCCTGCCGCTCTACAGCA 645
D 640 GCTCCGTGTGAGCAACAGCCCTTTGAGTGTGAGAAAGCCCTGCCGCTCTACAGCA 699
Q 646 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGCTGAAGA 705
D 700 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGCTGAAGA 759
Q 706 CCCCTCCCTCAACCTGATATCTGTGCACTGTGTCGGGACCCCGGGCGCTGTTCGTT 765
D 760 CCCCTCCCTCAACCTGATATCTGTGCACTGTGTCGGGACCCCGGGCGCTGTTCGTT 819
Q 766 CCGAGAACGCAAAAGGAGATCTCATGTGATGACAGTCCGATTTGTGAGGCGACATGA 825
D 820 CCGAGAACGCAAAAGGAGATCTCATGTGATGACAGTCCGATTTGTGAGGCGACATGA 879
Q 826 GCAAAACCTCAAGAGAGAGCAACCCCTACTATGTGATGACAGTCTGCCAAAGCCA 885
D 880 GCAAAACCTCAAGAGAGAGCAACCCCTACTATGTGATGACAGTCTGCCAAAGCCA 939
Q 886 GCTGAGATCTCAAGAGACATCCAGTCTTGGCCAAAGCCCTCAGAGAACGCTACTGCT 945
D 940 GCTGAGATCTCAAGAGACATCCAGTCTTGGCCAAAGCCCTCAGAGAACGCTACTGCT 999
Q 946 TTGTCCTATGAGAGCTGGCTGTGAGCCCTGTGGCCAGACTTCCGAAATGATGATAT 1005
D 1000 TTGTCCTATGAGAGCTGGCTGTGAGCCCTGTGGCCAGACTTCCGAAATGATGATAT 1059
Q 1006 CGTGGATGGAATTTTGGCCCATCTTCAAGCTGGTGCATATCAATCACCAGAGCAA 1065
D 1060 CGTGGATGGAATTTTGGCCCATCTTCAAGCTGGTGCATATCAATCACCAGAGCAA 1119
Q 1066 GGGCATGAGGTGACCAAGCTTTCACACAAATGCGAGGATGCCCTTAATGTCTCCAGGC 1125
D 1120 GGGCATGAGGTGACCAAGCTTTCACACAAATGCGAGGATGCCCTTAATGTCTCCAGGC 1179
Q 1126 TTGGGCTGTGCTTTTGGCCCATATGAAGTTTCTCACTTCAAAAGCCCTGTGGCATGC 1185
D 1180 TTGGGCTGTGCTTTTGGCCCATATGAAGTTTCTCACTTCAAAAGCCCTGTGGCATGC 1239
Q 1186 CATGAATTTGCTGGGTACCGCCAGCTCAGATCTGAACAAGACAGAAACCTGTTCCT 1245
D 1240 CATGAATTTGCTGGGTACCGCCAGCTCAGATCTGAACAAGACAGAAACCTGTTCCT 1299
Q 1246 GGATCTTCTGTCTACCTGACATGCTGCCCTGAGCAAAATCCACTAAGAGGGTTGAGAAGCT 1305
D 1300 GGATCTTCTGTCTACCTGACATGCTGCCCTGAGCAAAATCCACTAAGAGGGTTGAGAAGCT 1359
Q 1306 TTGTCACACTGTGTGACAGCTTCAAGTCTTCTGAAATGCTTCTGAGCTTGCCTACA 1365
D 1360 TTGTCACACTGTGTGACAGCTTCAAGTCTTCTGAAATGCTTCTGAGCTTGCCTACA 1419
Q 1366 TTCTGAGCTTAACTACATGCTGTGGGTATCAGACTGAGTGTGATTTGTCCACAG 1425
D 1420 TTCTGAGCTTAACTACATGCTGTGGGTATCAGACTGAGTGTGATTTGTCCACAG 1479
Q 1426 TTCTCAAGAGAGAACTTTTGTGTCAGTGTGTCTAAGAAACAGAGATGGGGAACCT 1485
D 1480 TTCTCAAGAGAGAACTTTTGTGTCAGTGTGTCTAAGAAACAGAGATGGGGAACCT 1539
Q 1486 TTATGAGAGAGAGCAATCCACAGAGTGAACAGGGTATTTGCTTCTTCTTCTTATCT 1545
D 1540 TTATGAGAGAGAGCAATCCACAGAGTGAACAGGGTATTTGCTTCTTCTTCTTATCT 1599
Q 1546 TTCTGTCTGAGAGACTTCAAGAGATTTGTGCTGTGAGGCTATTAAGACAGACAGT 1605
D 1600 TTCTGTCTGAGAGACTTCAAGAGATTTGTGCTGTGAGGCTATTAAGACAGACAGT 1659

Q 1606 ATCAGTGAATGATCCATTAACCTCCCTGTCACATCTTGGCCAAATGGGAATGATCT 1665
D 1660 ATCAGTGAATGATCCATTAACCTCCCTGTCACATCTTGGCCAAATGGGAATGATCT 1719
Q 1666 TTACCAAGAGCTCAACACCACTTTTCCACAGAGATGC 1703
D 1720 TTACCAAGAGCTCAACACCACTTTTCCACAGAGATGC 1757

RESULT 2
US-09-190-911-2
; Sequence 2, Application US/09190911
; Patent No. 6363365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmrich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/09/190,911
; CURRENT FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens
US-09-190-911-2

Query Match 75.7%; Score 1547; DB 4; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Q 106 CAAGTCTTCCACTTACAGACAAATGCTACTGCTTAAAAAATGAAGCTCTGCTTTCT 165
D 160 CAAGTCTTCCACTTACAGACAAATGCTACTGCTTAAAAAATGAAGCTCTGCTTTCT 219
Q 166 GCTTCCAGATGGCATCTTGGCTATCTTCCACATGATACAGCAACATCAAGCTC 225
D 220 GCTTCCAGATGGCATCTTGGCTATCTTCCACATGATACAGCAACATCAAGCTC 279
Q 226 CCGTCTATGAAGCAGACGCCAGCGCATGACAGTGTGCTGCTCTCTGAGGCTC 285
D 280 CCGTCTATGAAGCAGACGCCAGCGCATGACAGTGTGCTGCTCTCTGAGGCTC 339
Q 286 TGCTCTTCTTTTGTGGGAGCTTTTGGGAGCAACCCAGATGTTTCTACCTGATGA 345
D 340 TGCTCTTCTTTTGTGGGAGCTTTTGGGAGCAACCCAGATGTTTCTACCTGATGA 399
Q 346 GCGCGCTGGCAGCTGTGATGACATTCAGAGCAGACACCGCTGATGCTGACATGC 405
D 400 GCGCGCTGGCAGCTGTGATGACATTCAGAGCAGACACCGCTGATGCTGACATGC 459
Q 406 TTGTGCGGATCTGATACGGGCGCTTCTTGTGTCAGATGAGGCTTTTGATGCTACAT 465
D 460 TTGTGCGGATCTGATACGGGCGCTTCTTGTGTCAGATGAGGCTTTTGATGCTACAT 519
Q 466 GGAACCTGTGCCCGAGAGACATGCAAGCTCTTTCAGTGGAGAACAGCCGGGCTGTG 525
D 520 GGAACCTGTGCCCGAGAGACATGCAAGCTCTTTCAGTGGAGAACAGCCGGGCTGTG 579
Q 526 TTCTGACCTGCTGTGACATCATCCACAAGATGAATATCCCGGGCTCAGTGCAG 585
D 580 TTCTGACCTGCTGTGACATCATCCACAAGATGAATATCCCGGGCTCAGTGCAG 639
Q 586 GCTCTGTGAGCAACAGCCCTTTGAGTGTGAGAAAGCCCTGCCGCTCTACAGCA 645
D 640 GCTCTGTGAGCAACAGCCCTTTGAGTGTGAGAAAGCCCTGCCGCTCTACAGCA 699

QY	646	CGTGTGCTCAAGGAGGTGGCGCTTCTTCAACCTGCAGTCCCTTCAACCCGCTGTGAAGA	705
QY	700	CTGTGTCTCAAGGAGGTGGCGCTTCTTCAACCTGCAGTCCCTTCAACCCGCTGTGAAGA	759
QY	706	CCCTCCCTCAACCTGCATATCTGTCACACTGGTCCGGGACCCCGGGCGCTGTTCCGTTTC	765
DB	760	CCCTCCCTCAACCTGCATATCTGTCACACTGGTCCGGGACCCCGGGCGCTGTTCCGTTTC	819
QY	766	CCGAGAACGCACAAAGGAGATCTCATGTATTGACAGTGGCATTTGTGATGGGGCAGCATGA	825
DB	820	CCGAGAACGCACAAAGGAGATCTCATGTATTGACAGTGGCATTTGTGATGGGGCAGCATGA	879
QY	826	GCAAAAACCTCAGAAGAGGAGACCAACCCCTACTATGTGATGGAGGCATCTGCCAAAGCCA	885
DB	880	GCAGAAACCTCAAGAAGAGGAGACCAACCCCTACTATGTGATGGAGGCATCTGCCAAAGCCA	939
QY	886	GCTGGAGATCTTACAAGACCATCCAGTCTCTTGCCCAAGGCGCTGCAGAGAGCTACTGCT	945
DB	940	GCTGGAGATCTTACAAGACCATCCAGTCTCTTGCCCAAGGCGCTGCAGAGAGCTACTGCT	999
QY	946	TGTGCGCTATAGAGAGCTGGGCTGCAGGCGCTGTGGCCCGCAGACTTCCCAATGTATGAATT	1005
DB	1000	TGTGCGCTATAGAGAGCTGGGCTGCAGGCGCTGTGGCCCGCAGACTTCCCAATGTATGAATT	1035
QY	1006	CGTGGGATTTGGAATTTCTTGGCCCATCTTTCAGACCTGGGTGCATAACATCACCCGAGGCAA	1065
DB	1060	CGTGGGATTTGGAATTTCTTGGCCCATCTTTCAGACCTGGGTGCATAACATCACCCGAGGCAA	1119
QY	1066	GGGCAATGGGTATACCAACGCTTTCACACAAATGGCCAGGGATGCCCTATATGTCTCCAGGC	1125
DB	1120	GGGCAATGGGTATACCAACGCTTTCACACAAATGGCCAGGGATGCCCTATATGTCTCCAGGC	1179
QY	1126	TTGGCGCTGTCTTTCGCCCTATGAAAAGGTTTCTCGACTTCAGAAAGGCTGTGGCGATGC	1185
DB	1180	TTGGCGCTGTCTTTCGCCCTATGAAAAGGTTTCTCGACTTCAGAAAGGCTGTGGCGATGC	1239
QY	1186	CATGAATTTGGTGGGCTACCGCCACAGTACGATCTGAAACAAAGACAGAAACCTGTCTCT	1245
DB	1240	CATGAATTTGGTGGGCTACCGCCACAGTACGATCTGAAACAAAGACAGAAACCTGTCTCT	1299
QY	1246	GGATCTTCTGTCTACTGAGTGTGTCCTCGACAAATCCACTAAGAGGTTGAGAAGCTT	1305
DB	1300	GGATCTTCTGTCTACTGAGTGTGTCCTCGACAAATCCACTAAGAGGTTGAGAAGCTT	1355
QY	1306	TGCTGACCACTGTGTCTACAGCTCAAGTCACTTCTCTGATGATCTTGAGCTTGCTGCTACA	1365
DB	1360	TGCTGACCACTGTGTCTACAGCTCAAGTCACTTCTCTGATGATCTTGAGCTTGCTGCTACA	1419
QY	1366	TCTCTGAGACCTTAACTACATGTCGTGGGTATCCACACGAGTGGAGTGTGTCCACAGC	1425
DB	1420	TCTCTGAGACCTTAACTACATGTCGTGGGTATCCACACGAGTGGAGTGTGTGTCCACAGC	1479
QY	1426	TGCTCAAGCAGAAAGACTTTGTGTCTCATGCTTGTGTCTAGAAAAACAGACATGGGGAACTT	1485
DB	1480	TGCTCAAGCAGAAAGACTTTGTGTCTCATGCTTGTGTCTAGAAAAACAGACATGGGGAACTT	1539
QY	1486	TATGTGAGACACATCCACCACTGAAACAGGGTATTTGCTCTTCTTTCTTGTGATCT	1545
DB	1540	TATGTGAGACACATCCACCACTGAAACAGGGTATTTGCTCTTCTTTCTTGTGATCT	1599
QY	1546	TCTCTGTCTGGGACAGCTTCAGAGACTTTGTGGCGGAGGGCTCTTTAAGCAGACACAGT	1605
DB	1600	TCTCTGTCTGGGACAGCTTCAGAGACTTTGTGGCGGAGGGCTCTTTAAGCAGACACAGT	1659
QY	1606	ATCAGTGAATTTGATTCATAAACCTCCCTGTCCACATCTTCCCAATGGGAATGAGATCT	1719
DB	1666	TTCAACCAAGAGCTCACACAGATTTTCCACAGAGATGC 1703	
QY	1700	TTCAACCAAGAGCTCACACAGATTTTCCACAGAGATGC 1757	

```

RESULT 3
US-09-786-240-26
; Sequence 26, Application US/09786240
; Patent No. 6558935
; GENERAL INFORMATION:
; APPLICANT: INCYTE PHARMACEUTICALS, INC.
; APPLICANT: TANG, Y. TOM
; APPLICANT: CORLEY, Neil C.
; APPLICANT: GUEGLER, Karl J.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LAL, Preeti
; APPLICANT: YUE, Henry
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: AZIMZAI, Valda
; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
; FILE REFERENCE: PF-0592 PCT
; CURRENT APPLICATION NUMBER: US/09/786,240
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 60/133,6
; PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1999-05-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 26
; LENGTH: 2065
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: incyte ID No. 6558935 2617407CB1
US-09-786-240-26

```

Query Match	56.4%	Score 1153:	DB 4:	Length 2065:
Best Local Similarity	99.6%	Pred. No. 0:		
Matches 193:	Conservative	0:	Mismatches	4: Indels
			Gaps	2:
QY	106	CAAGTCTTCACCTTCAGACACAATGCTACTGCTAAAAAATGAAGCTCTGCTGTTCT	165	
Db	152	CAAGTCTTCACCTTCAGACACAATGCTACTGCTAAAAAATGAAGCTCTGCTGTTCT	211	
QY	166	GGTTTCCAGATGGGCATCTTGGGCTCATCTTCTCCACATGTAACGCCAACATCAGCTC	225	
Db	212	GGTTTCCAGATGGGCATCTTGGGCTCATCTTCTCCACATGTAACGCCAACATCAGCTC	271	
QY	226	CCTGTCTATGAAGGCACAGCCCGAGCCGATGCACGCTGCTGTTCTGCTTCCCTGGAGCTC	285	
Db	272	CCTGTCTATGAAGGCACAGCCCGAGCCGATGCACGCTGCTGTTCTGCTTCCCTGGAGCTC	331	
QY	286	TGGCTCTCTTTTGGGGGAGCCTTTTGGGAGAGCACCACATGTTTCTACCTGATGGA	345	
Db	332	TGGCTCTCTTTTGGGGGAGCCTTTTGGGAGAGCACCACATGTTTCTACCTGATGGA	391	
QY	346	GCCCGCCTGGGACAGCTGTGATGACCTTCAAGCAGAGACCCGCTTGATGCTGCACATGGC	405	
Db	392	GCCCGCCTGGGACAGCTGTGATGACCTTCAAGCAGAGACCCGCTTGATGCTGCACATGGC	451	
QY	406	TGTGGGGATCTGATACGGGCGCTTCTTGTGGCACAATGAGCGCTTTGATGCTCAAT	465	
Db	452	TGTGGGGATCTGATACGGGCGCTTCTTGTGGCACAATGAGCGCTTTGATGCTCAAT	511	
QY	466	GGAACCTGTCCTCCGGAGACAGTCACGCGCTCTTTCAGTGGGAGAACACCGGCGCCTGTG	525	
Db	512	GGAACCTGTCCTCCGGAGACAGTCACGCGCTCTTTCAGTGGGAGAACACCGGCGCCTGTG	571	
QY	526	TTTGTACACTGCGCTGTGACATCATCCACAAGATGAATTCATCCCGGGGCTCACTGCAG	585	
Db	572	TTTGTACACTGCGCTGTGACATCATCCACAAGATG - AATTCATCCCGGGGCTCACTGCAG	630	
QY	586	GCTCCTGTGCAGTAAACAGCCCTTGTG - AGGTGGTGGAGAAAGCCGCGCTCTACAGCC	644	
Db	631	GCTCCTGTGCAGTAAACAGCCCTTGTGAAGTTGTGGAGAAAGCCGCGCTCTACAGCC	690	
QY	645	ACGTGTGTCTCAAGAGGTGCGCTTCTTCAACTGTACGTCCCTTACCCGCTGCTGAAG	704	

```
Db 691 AGTGTGCTCAAGAGGTCGCTTCTTAACCTGACGTCCTTACCCGCTGCTGAAG 750
QY 705 ACCCTCCCTCAACCTGCATATCTGTCACACCTGGTCCGGAGCCCGCGGCGTTCCTT 764
Db 751 ACCCTCCCTCAACCTGCATATCTGTCACACCTGGTCCGGAGCCCGCGGCGTTCCTT 810
QY 765 CCCGAGAACGCAAAAGGAGATCTCATATTTGACAGTGTGCATTTGATGGGCGACATG 824
Db 811 CCCGAGAACGCAAAAGGAGATCTCATATTTGACAGTGTGCATTTGATGGGCGACATG 870
QY 825 AACAAAACCTCAAGAGAGGAGCAACCCCTACTATGTATGAGGATCATCTGCCAAGCC 884
Db 871 AACAGAACTCAAGAGAGGAGCAACCCCTACTATGTATGAGGATCATCTGCCAAGCC 930
QY 885 AGCTGAGATCTCAAGAGACCATCATCTCTGCCCCAAGGCGCTGCAGAGACCTTACTC 944
Db 931 AGCTGAGATCTCAAGAGACCATCATCTCTGCCCCAAGGCGCTGCAGAGACCTTACTC 990
QY 945 TTGTGGCTATAGAGACCTGCTCCAGCCCTGTGGCCCGACAGCTTCCGAAATGATGAT 1004
Db 991 TTGTGGCTATAGAGACCTGCTCCAGCCCTGTGGCCCGACAGCTTCCGAAATGATGAT 1050
QY 1005 TCGTGGGATTTGAAATTTCTTGCCCACTTACAGACCTGGGTGATATACATCACCCGAG 1064
Db 1051 TCGTGGGATTTGAAATTTCTTGCCCACTTACAGACCTGGGTGATATACATCACCCGAG 1110
QY 1065 AGGGCATGGGTGATGACCAAGCTTTCACACAATGCCAGGAGTCCCTTAATGTCTCCAG 1124
Db 1111 AGGGCATGGGTGATGACCAAGCTTTCACACAATGCCAGGAGTCCCTTAATGTCTCCAG 1170
QY 1125 CTGGCGCTGTGCTTTGCCCTATGAAAAGTTTCTGATTTGAGAAAGCTGTGGGATG 1184
Db 1171 CTGGCGCTGTGCTTTGCCCTATGAAAAGTTTCTGATTTGAGAAAGCTGTGGGATG 1230
QY 1185 CCATGATTTGCTGGGCTACCCGACAGTGCAGATCTGAACAAGAGAGAAACCTGTTC 1244
Db 1231 CCATGATTTGCTGGGCTACCCGACAGTGCAGATCTGAACAAGAGAGAAACCTGTTC 1290
QY 1245 TGGATCTTCTGTCTACCTGAGCTGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAGGCT 1304
Db 1291 TGGATCTTCTGTCTACCTGAGCTGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAGGCT 1350
QY 1305 TTGCTGCGACGCTGGGTGACGCTCACTTCTCTGATGCTTCTGAGCCCTGCTGAC 1364
Db 1351 TTGCTGCGACGCTGGGTGACGCTCACTTCTCTGATGCTTCTGAGCCCTGCTGAC 1410
QY 1365 ATCTGTGACCTTAATACATGTCTGTGGGTATCAACATGATGTGATGTGTCCACAC 1424
Db 1411 ATCTGTGACCTTAATACATGTCTGTGGGTATCAACATGATGTGATGTGTCCACAC 1470
QY 1425 GTGCTCAAGCAGAGAGACTTTGTGTCCATGCTGTGTCTAGAAAACAGACTGGGAGCC 1484
Db 1471 GTGCTCAAGCAGAGAGACTTTGTGTCCATGCTGTGTCTAGAAAACAGACTGGGAGCC 1530
QY 1485 TATGTGACGACACATCCACAGAGGATGTGCTTCTTCTTCTTCTTCTTCTGATC 1544
Db 1531 TATGTGACGACACATCCACAGAGGATGTGCTTCTTCTTCTTCTTCTTCTTCTGATC 1590
QY 1545 TTCTGTCTGGGACAGCTTCAAGAGCTTTGTGGCTGAGGCGCTATTAAAGCAGACAG 1604
Db 1591 TTCTGTCTGGGACAGCTTCAAGAGCTTTGTGGCGTGGAGGCGCTATTAAAGCAGACAG 1650
QY 1605 TATCAGTGAATTTGATCCATAAAGCTCCCTGTCCACATTTGCCCCAATGGGAGATGATC 1664
Db 1651 TATCAGTGAATTTGATCCATAAAGCTCCCTGTCCACATTTGCCCCAATGGGAGATGATC 1710
QY 1665 TTTCACAAAGAGCTCACAGCATTTTTCACAGAGATGC 1703
Db 1711 TTTCACAAAGAGCTCACAGCATTTTTCACAGAGATGC 1749
```

RESULT 4
US-09-045-284A-7

```
Sequence 7, Application US/09045284A
; Patent No. 6265192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107051
; CURRENT APPLICATION NUMBER: US/09/045,284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-045-284A-7

Query Match 1.8%; Score 37; DB 3; Length 37;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 830 AAACCTCAAGAGAGGAGCAACCCCTACTATGTGATGC 866
1 AAACCTCAAGAGAGGAGCAACCCCTACTATGTGATGC 37

RESULT 5
US-09-190-911-7
; Sequence 7, Application US/09190911
; Patent No. 6363365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Fangermann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107051
; CURRENT APPLICATION NUMBER: US/09/190,911
; CURRENT FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: DNA
; ORGANISM: H. sapiens
US-09-190-911-7

Query Match 1.8%; Score 37; DB 4; Length 37;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 830 AAACCTCAAGAGAGGAGCAACCCCTACTATGTGATGC 866
1 AAACCTCAAGAGAGGAGCAACCCCTACTATGTGATGC 37

RESULT 6
US-08-992-334-1
; Sequence 1, Application US/08992334
; Patent No. 5919678
; GENERAL INFORMATION:
; APPLICANT: Gruss, Alexandra
; APPLICANT: Maguin, Emmanuelle
; TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
; TITLE OF INVENTION: PLASMID
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christie Parker & Hale, LLP
; STREET: 350 West Colorado Boulevard, Suite 500
; CITY: Pasadena
```


STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR FR92/03034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3792 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: YES
IMMEDIATE SOURCE:
CLONE: pc-host4
US-08-992-334-1

Query Match 0.9%; Score 19; DB 2; Length 3792;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCCTGACAAAGAGCAAG 1839
1818 TCCTGACAAAGAGCAAG 1836

RESULT 7
US-08-302-752-1
Sequence 1, Application US/08302752
Patent No. 6025190
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: THERMOSENSIBLE PLASMID
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3792 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-1

Query Match 0.9%; Score 19; DB 3; Length 3792;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCCTGACAAAGAGCAAG 1839
1818 TCCTGACAAAGAGCAAG 1836

RESULT 8
US-08-992-334-2
Sequence 2, Application US/08992334
Patent No. 5919678
GENERAL INFORMATION:
APPLICANT: Gruss, Alexandra
APPLICANT: Maguin, Emmanuelle
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
MOLECULE TYPE: PLASMID
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christie Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR FR92/03034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5234 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-992-334-2

Query Match 0.9%; Score 19; DB 2; Length 5234;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1821 TCTGTGACAAAGAGCAG 1839
DB 3260 TCTGTGACAAAGAGCAG 3278

RESULT 9

US-08-302-752-2
Sequence 2, Application US/08302752
Patent No. 6025190
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: THERMOSENSIBLE PLASMID
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5234 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-2

Query Match
Best Local Similarity 100.0%, Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1821 TCTGTGACAAAGAGCAG 1839
DB 3260 TCTGTGACAAAGAGCAG 3278

RESULT 10
US-08-992-334-3
Sequence 3, Application US/08992334
Patent No. 5919678
GENERAL INFORMATION:
APPLICANT: Gruss, Alexandra
APPLICANT: Maguin, Emmanuelle
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
TITLE OF INVENTION: PLASMID
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christle Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/03034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6722 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-992-334-3

Query Match
Best Local Similarity 100.0%, Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1821 TCTGTGACAAAGAGCAG 1839
DB 4748 TCTGTGACAAAGAGCAG 4766

RESULT 11
US-08-302-752-3
Sequence 3, Application US/08302752
Patent No. 6025190
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: THERMOSENSIBLE PLASMID
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6722 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-3

Query Match
Best Local Similarity 100.0%, Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1821 TCTGTGACAAAGAGCAG 1839
DB 4748 TCTGTGACAAAGAGCAG 4766

```
RESULT 12
US-08-778-494B-112
: Sequence 112, Application US/08778494B
: Patent No. 5962272
: GENERAL INFORMATION:
: APPLICANT: Chenchik, Alex
: APPLICANT: Zhu, York
: APPLICANT: Diachenko, Luda
: APPLICANT: Siebert, Paul
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR FULL-LENGTH CDNA
: TITLE OF INVENTION: CLONING
: NUMBER OF SEQUENCES: 114
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: Florida
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION NUMBER: US/08/778,494B
: APPLICATION NUMBER: US/08/778,494B
: FILING DATE: 03-JAN-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/582,562
: FILING DATE: 03-JAN-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Pace, Doran R.
: REGISTRATION NUMBER: 38,261
: REFERENCE/DOCKET NUMBER: CL-7C1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (352) 375-8100
: TELEFAX: (352) 372-5800
: INFORMATION FOR SEQ ID NO: 112:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 23 bases
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (synthetic)
: US-08-778-494B-112

Query Match          0.9%; Score 18; DB 2; Length 23;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      21 CGGAAGACGACAGAGGG 38
DB      6 CGGAAGACGACAGAGGG 23

RESULT 13
US-09-045-284A-8/c
: Sequence 8, Application US/09045284A
: Patent No. 6265192
: GENERAL INFORMATION:
: APPLICANT: Bistrup, Annette
: APPLICANT: Rosen, Steven D.
: APPLICANT: Hemmerich, Stefan
: TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
: FILE REFERENCE: 6510-107U01
: CURRENT APPLICATION NUMBER: US/09/045,284A
: CURRENT FILING DATE: 1998-03-20
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 8
: LENGTH: 47
```

```
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-045-284A-8

Query Match          0.9%; Score 18; DB 3; Length 47;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1248 ATCTTCGTCTACTCGA 1265
DB      47 ATCTTCGTCTACTCGA 30

RESULT 14
US-09-190-911-8/c
: Sequence 8, Application US/09190911
: Patent No. 6365365
: GENERAL INFORMATION:
: APPLICANT: Bistrup, Annette
: APPLICANT: Rosen, Steven D.
: APPLICANT: Tangemann, Kirsten
: APPLICANT: Hemmerich, Stefan
: TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
: FILE REFERENCE: 6510-107CIP
: CURRENT APPLICATION NUMBER: US/09/190,911
: CURRENT FILING DATE: 1998-11-12
: EARLIER APPLICATION NUMBER: 09/045,284
: EARLIER FILING DATE: 1998-03-20
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 8
: LENGTH: 47
: TYPE: DNA
: ORGANISM: H. sapiens
US-09-190-911-8

Query Match          0.9%; Score 18; DB 4; Length 47;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1248 ATCTTCGTCTACTCGA 1265
DB      47 ATCTTCGTCTACTCGA 30

RESULT 15
US-09-439-313-320
: Sequence 320, Application US/09439313
: Patent No. 6329505
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: APPLICANT: Dillon, Davin C.
: APPLICANT: Mitcham, Jennifer L.
: APPLICANT: Harlocker, Susan Louise
: APPLICANT: Jiang Yuqul
: APPLICANT: Reed, Steven G.
: APPLICANT: Kalos, Michael
: APPLICANT: Fanger, Gary
: APPLICANT: Reiter, Mark
: APPLICANT: Solk, John
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
: FILE REFERENCE: 210121.427C9
: CURRENT APPLICATION NUMBER: US/09/439,313
: CURRENT FILING DATE: 1999-11-12
: NUMBER OF SEQ ID NOS: 575
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 320
: LENGTH: 150
: TYPE: DNA
: ORGANISM: Homo sapien
US-09-439-313-320
```

Query Match 0.98; Score 18; DB 4; Length 150;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAATTCATTGTGTGG 18
|||||
Db 31 GAATTCATTGTGTGG 48

Search completed: August 14, 2003, 21:19:26
Job time : 157 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 17:40:11 ; Search time 3887 Seconds
(without alignments)
12774.388 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043

Sequence: 1 gaattccattggttggtgta.....tgggatacctaataaaaaa 2043

Scoring table:

Gapop 60.0 , Gapept 60.0

Searched: 22781392 seqs, 1215238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562670

Minimum DB seq length: 10

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estda:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_liv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gssl:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	685	33.5	954	BI823850 603039012
2	579	28.3	583	BK490456 Homo sapi
3	534	26.1	668	AL709927 DKF2P8860
4	479	23.4	571	BM129080 1f17c04.y

Result No.	Score	Query Length	ID	Description
5	322	15.8	593	BM969292 UI-CF-ENO
6	290	14.2	680	BF878439 MRO-EMO10
7	288	14.1	669	B0684395 UI-CF-ENO
8	218	10.7	553	BM128370 1f13d12.x
9	218	10.7	553	BM128370 1f13d12.x
10	107	5.2	362	AM002418 wu61903.x
11	105	5.1	417	AM572510 xq18911.x
12	101	4.9	358	BF056840 7x10d12.x
13	74	3.6	419	AM572390 xuo9a06.x
14	59	2.9	525	BF197521 7684d08.x
15	59	2.9	620	AM245312 K0722H07-
16	59	2.9	695	BE857538 7601a08.x
17	59	2.9	735	BE858652 7601a09.x
18	55	2.7	721	AO741911 HS-5569_B
19	47	2.3	436	AM081348 xc41b06.x
20	27	1.3	507	BM246681 K0741E04-
21	27	1.3	571	BM245312 K0722H07-
22	27	1.3	695	BE854671 602831875
23	27	1.3	783	BG963298 602827716
24	27	1.3	852	BG966340 602832826
25	27	1.3	1923	AK009113 MNS MUSCU
26	27	1.1	362	BE340900 EST344988
27	23	1.1	397	AI345721 t684d04.x
28	23	1.1	687	CB167190 GWM602702
29	23	1.1	692	BY711808 BY711808
30	23	1.1	834	B0884877 AGENCOURT
31	22	1.1	148	AI345073 t652d07.x
32	22	1.1	160	B0265877 NISC-ET09
33	22	1.1	167	AI253971 qv54a05.x
34	22	1.1	171	AI733948 qv54a05.x
35	22	1.1	171	AI246069 qv52f03.x
36	22	1.1	187	AI251238 qv55905.x
37	22	1.1	187	AI792899 qv55905.y
38	22	1.1	191	AI251979 qv57d08.x
39	22	1.1	196	AI792857 qv54a05.y
40	22	1.1	253	AA076760 7801F06 C
41	22	1.1	290	AI345395 t682b03.x
42	22	1.1	361	AI311365 t647h05.x
43	22	1.1	493	BF869376 t63-EP011
44	22	1.1	501	BH287165 CH230-113
45	22	1.1	516	AA689056 v806c06.r

ALIGNMENTS

RESULT 1
BI823850 954 bp mRNA linear EST 04-OCT-2001
LOCUS 603039012F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5179926 5',
DEFINITION mRNA sequence.
ACCESSION BI823850
VERSION BI823850.1 GI:15935400
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE NIH-MGC http://mgc.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: L1AM11448 row: h column: 03
High quality sequence stop: 856.

QY 281 CGCTGCGCTCTTCTTTTGTGGGGAGAGCTTTTGGGAGACCCAGATGTTTCTACTG 340
 Db 241 CGCTGCGCTCTTCTTTTGTGGGGAGAGCTTTTGGGAGACCCAGATGTTTCTACTG 300
 QY 341 ATGAGAGCCGCTTGGGAGAGCTTGTGAGTACCTTCAAGCAGAGACCCGCTGATGCTGAC 400
 Db 301 ATGAGAGCCGCTTGGGAGAGCTTGTGAGTACCTTCAAGCAGAGACCCGCTGATGCTGAC 360
 QY 401 ATGAGAGCTTGGGAGATGATACGAGGCGCTTCTTGTGGAGCATGAGCTTTTGAAGCC 460
 Db 361 ATGAGCTGTGGGAGATGATACGAGGCGCTTCTTGTGGAGCATGAGCTTTTGAAGCC 420
 QY 461 TACATGGAACCTGTGCTCCCGGAGAGACATCCAGCTTTTCAAGTGGAGAACAGCCGGCC 520
 Db 421 TACATGGAACCTGTGCTCCCGGAGAGACATCCAGCTTTTCAAGTGGAGAACAGCCGGCC 480
 QY 521 CTGTGTTTCTGACCTGCTGCTGTGACATATCCACAGATGAATATCCCGGGGCTCAC 580
 Db 481 CTGTGTTTCTGACCTGCTGCTGTGACATATCCACAGATGAATATATCCCGGGGCTCAC 540
 QY 581 TGCAGGCTCCTGTGACATCAACAGCCCTTGAAGTGTG 619
 Db 541 TGCAGGCTCCTGTGACATCAACAGCCCTTGAAGTGTG 579

RESULT 3
 AL709927 668 bp mRNA linear EST 12-JUN-2003
 LOCUS DKFZp68602364.F1 686 (synonym: hlc3) Homo sapiens cDNA clone
 AL709927 DKFZp68602364.5, mRNA sequence.
 AL709927.1 GI:19693282

ACCESSION AL709927
 VERSION AL709927.1
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 668)
 Authors: Bahr, A., Lauber, J., Mewes, H.W., Well, B., Amid, C., Osanger, A., Fobo, G., Han, M. and Wiemann, S.
 Title: EST (Bahr, A., Lauber, J., Mewes, H.W., Well, B., et al.)
 Journal: Unpublished
 Comment: Contact: Bahr A
 MFS

Ingolstaedter Landstr. 1, D-85764 Neuherberg, Germany
 This is the 5' sequence of the clone insert
 clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de;
 sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
 consortium of the German Genome Project.
 No s1 sequence available.
 This clone (DKFZp68602364) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 Location/Qualifiers

FEATURES

source
 1. 668
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFZp68602364"
 /tissue_type="human skeletal muscle"
 /dev_stage="adult"
 /lab_host="DH10s"
 /clone_id="668 (synonym: hlc3)"
 /note="vector: pTriblex2; Site_1: SfiIA; Site_2: SfiIB;
 cDNA-collection"

BASE COUNT 132 a 202 c 178 g 154 t 2 others
 ORIGIN

Query Match 26.1%; Score 534; DB 9; Length 668;
 Best Local Similarity 100.0%; Pred. No. 2.3e-209;
 Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 AGAGAGAAAAAGCGATGCGCCGCGCTAGACATGAGCTCTCTAAAGCAGAGGAGAGCC 99
 Db 6 AGAGAGAAAAAGCGATGCGCCGCGCTAGACATGAGCTCTCTAAAGCAGAGGAGAGCC 65
 QY 100 AAGCCAGAGGCTTTCACATTCAGACATGATGCTCTCTAAAGGAGAGCTCTGCT 159
 Db 66 AAGCCAGAGGCTTTCACATTCAGACATGATGCTCTCTAAAGGAGAGCTCTGCT 125
 QY 160 GTTTCGTTTCCAGATGCGCATCTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 219
 Db 126 GTTTCGTTTCCAGATGCGCATCTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 185
 QY 220 CAGTCCCTGTCTATGAAGGACAGCCGAGCGCATGCTGCTGCTCTCTCTCTCTCT 279
 Db 186 CAGTCCCTGTCTATGAAGGACAGCCGAGCGCATGCTGCTGCTCTCTCTCTCTCT 245
 QY 280 GCGCTGCGCTCTCTTTTGTGGGAGAGCTTTTGGGAGAGCCAGATGTTTCTACT 339
 Db 246 GCGCTGCGCTCTCTTTTGTGGGAGAGCTTTTGGGAGAGCCAGATGTTTCTACT 305
 QY 340 GATGAGAGCCGCGCTGAGAGCTGATGATGATGATGATGATGATGATGATGATGAT 399
 Db 306 GATGAGAGCCGCGCTGAGAGCTGATGATGATGATGATGATGATGATGATGATGAT 365
 QY 400 CATGCTGTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 459
 Db 366 CATGCTGTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 425
 QY 460 CTACATGGAACCTGTGCTCCCGGAGAGACATCCAGCTTTTCAAGTGGAGAACAGCCGGCC 519
 Db 426 CTACATGGAACCTGTGCTCCCGGAGAGACATCCAGCTTTTCAAGTGGAGAACAGCCGGCC 485
 QY 520 CCGTGTTCGACCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 573
 Db 486 CCGTGTTCGACCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 539

RESULT 4
 BM129080 571 bp mRNA linear EST 12-MAR-2002
 LOCUS I117C04.Y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
 cDNA clone IMAGE:567665.5 similar to TR:0915R3 Q915R3
 N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE. ;, mRNA sequence.
 BM129080
 BM129080.1 GI:17123632

ACCESSION BM129080
 VERSION BM129080.1
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 571)
 Authors: Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, D., Blaisdell, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagaris, R., Williams, T., Jackson, Y. and Bowers, Y.
 Title: Endocrine Pancreas Consortium
 Journal: Unpublished
 Comment: Other ESTs: I117C04.x1
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept. of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohpc.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:
 Washington University Genome Sequencing Center For Information on
 obtaining a clone please contact: Juliana Brown
 (brown@fas.harvard.edu) This sequence now available from the IMAGE
 consortium, for clone orders contact: info@image.llnl.gov

The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p1773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)₁₈ tail. The sequence tag for this library is CTGCTCAGGT.

TAG_LIB-UI-CF-ENO
TAG_TISSUE-human Lung Epithelial Cell Lines untreated LPS
6hr to LPS 24h
TAG_SEQ-CTGCTCAGGT

BASE COUNT 191 a 138 c 147 g 193 t

Query Match 14.1%; Score 288; DB 13; Length 669;
Best Local Similarity 100.0%; Pred. No. 6.7e-108; Indels 0; Gaps 0;
Matches 288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1416 TGTCCACAGCTGCTCAGACAGAGAGACTTTGTGTCATGCTTGTCTAGAAACAGAC 1475
|||||
636 TGTCCACAGCTGCTCAGACAGAGAGACTTTGTGTCATGCTTGTCTAGAAACAGAC 577
|||||
1476 TGGGACACCTTATGTGAGACAGACATCCACAGTGAAGAGGATTCCTCTCTT 1535
|||||
576 TGGGACACCTTATGTGAGACAGACATCCACAGTGAAGAGGATTCCTCTCTCT 517
|||||
1536 TCTGTGATCTTCTGCTGGGAGAGACTTGTGCTGAGACTTTGGCCGAGGAGCTATTAAAC 1595
|||||
516 TCTGTGATCTTCTGCTGGGAGAGACTTGTGCTGAGACTTTGGCCGAGGAGCTATTAAAC 457
|||||
1556 ACGACACAGTATCAGTGAATTCATTAACCTCCCTGTCACATCTTGGCCCAATGGG 1655
|||||
456 ACGACACAGTATCAGTGAATTCATTAACCTCCCTGTCACATCTTGGCCCAATGGG 397
|||||
1656 GAATGATCTTTCACCAAGAGACTCACCAGCATTTTCCACAGAGATGC 1703
|||||
396 GAATGATCTTTCACCAAGAGACTCACCAGCATTTTCCACAGAGATGC 349
|||||

RESULT 8
BM128370/c 553 bp mRNA 1linear EST 12-MAR-2002
LOCUS lfi3d12.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
DEFINITION CDNA clone IMAGE:5676335 3', mRNA sequence.
ACCESSION BM128370
VERSION BM128370.1 GI:17122922
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Maria, M., Pape, D., Wylie, T., Martin, J., Blisstein, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagaris, R., Williams, T., Jackson, Y., and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished

TITLE
JOURNAL
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohpc.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For Information on

obtaining a clone please contact: Juliana Brown (brownjfas.harvard.edu) This sequence now available from the IMAGE consortium, for clone orders contact: info@image.llnl.gov
High quality sequence stop: 425.
Location/Qualifiers

FEATURES
source

1..553
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5676335"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/note="Organ: Pancreas; Vector: pSPORFL; Site 1: Not 1; Site 2: Sal 1; Starting library constructed using Superscript plasmid library kit (Life Technologies). cDNA made by oligo-dT priming. Size selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806. 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 163 a 111 c 127 g 152 t

Query Match 10.7%; Score 218; DB 12; Length 553;
Best Local Similarity 100.0%; Pred. No. 5.3e-79; Indels 0; Gaps 0;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1486 TATGAGACAGACATCCACAGTGAAGAGGATTCCTCTCTTCTTGATGCT 1545
|||||
553 TATGAGACAGACATCCACAGTGAAGAGGATTCCTCTCTTCTTGATGCT 494
|||||
1546 TCTGTGCTGGGAGAGCTTGTGCTGAGAGGCTTATTAAGACAGACAGT 1605
|||||
493 TCTGTGCTGGGAGAGCTTGTGCTGAGAGGCTTATTAAGACAGACAGT 434
|||||
1606 ATCAGTGAATTCATTAACCTCCCTGTCACATCTTGGCCCAATGGGATGATCT 1665
|||||
433 ATCAGTGAATTCATTAACCTCCCTGTCACATCTTGGCCCAATGGGATGATCT 374
|||||
1666 TTCACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 1703
|||||
373 TTCACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 336
|||||

RESULT 9
BM128831/c 553 bp mRNA 1linear EST 12-MAR-2002
LOCUS lfi7c04.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
DEFINITION CDNA clone IMAGE:5676655 3', mRNA sequence.
ACCESSION BM128831
VERSION BM128831.1 GI:17123383
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Maria, M., Pape, D., Wylie, T., Martin, J., Blisstein, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagaris, R., Williams, T., Jackson, Y., and Bowers, Y.
Endocrine Pancreas Consortium

TITLE
JOURNAL
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohpc.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For Information on

JOURNAL COMMENT Unpublished
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biochem.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center for information on
obtaining a clone please contact: Juliana Brown
(brownjefas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
High quality sequence stop: 425.

FEATURES

SOURCE

Location/Qualifiers

1..553

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:567655"

/sex="both"

/tissue_type="Islets of Langerhans"

/dev_stage="Adult"

/lab_host="DH10B"

/note="Organ: Pancreas; Vector: pSPORT1; Site: 1; Site_2: Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 163 a 111 c 127 g 152 t
ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 5.3e-79; Length 553;

Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1486 TATGTGAGCAGCAGATCCCAAGAGAGGATGCTCTTCTTTCTTGATCT 1545

DB 553 TATGTGAGCAGCAGATCCCAAGAGAGGATGCTCTTCTTTCTTGATCT 494

QY 1546 TCCGTCTGGGAGACTTCAGAGACTTGGGCTGAGGCGCTTTAAGCAGCAGACAGT 1605

DB 493 TCCGTCTGGGAGACTTCAGAGACTTGGGCTGAGGCGCTTTAAGCAGCAGACAGT 434

QY 1606 ATCAGTGAATTCATTAACCTCCCTGTCACATCTGGCCATGGGGAATGATCT 1665

DB 433 ATCAGTGAATTCATTAACCTCCCTGTCACATCTGGCCATGGGGAATGATCT 374

QY 1666 TTCACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 1703

DB 373 TTCACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 336

RESULT 10

AM002418/c 362 bp mRNA linear EST 27-OCT-1999

LOCUS w619g3.x1 NCI_CGAP_G66 Homo sapiens cDNA clone IMAGE:2524564 3'

DEFINITION mRNA sequence.

ACCESSION AM002418

VERSION AM002418.1 GI:5849334

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 362)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

R. Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bdrip/image/image.html

Seq primer: -40UP from Glibco.

FEATURES

SOURCE

Location/Qualifiers

1..362

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2524564"

/tissue_type="pooled germ cell tumors"

/lab_host="DH10B"

/clone_id="NCI_CGAP_G66"

/note="Vector: pYT3D-Pac (Pharmacia) with a modified

polylinker; Site: 1; Not I; Site 2: Eco RI; plasmid DNA

from the normalized library NCI CGAP GC4 was prepared, and

ss circles were made in vitro. Following HAP purification,

this DNA was used as tracer in a subtractive hybridization

reaction. The driver was PCR-amplified cDNAs from a pool

of 5,000 clones made from the same library (clone IDs

1257096-1258631, 1469064-1470983, and 1475592-1476743).

Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 108 a 69 c 77 g 108 t
ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 3.7e-33; Length 362;

Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1930 TGCATCTGTTAATCTAAATTTCCCAATTAAGTTCTGTTAATAGTCCCTTTATG 1989

DB 107 TGCATCTGTTAATCTAAATTTCCCAATTAAGTTCTGTTAATAGTCCCTTTATG 48

QY 1990 CTCTTAAATTTAGCAGTAATGTTCAATTTTATGGATCTTAA 2036

DB 47 CTCTTAAATTTAGCAGTAATGTTCAATTTTATGGATCTTAA 1

RESULT 11

AM572510 417 bp mRNA linear EST 13-MAR-2000

LOCUS xq18g11.x2 NCI_CGAP_Uc1 Homo sapiens cDNA clone IMAGE:2751044 3'

DEFINITION mRNA sequence.

ACCESSION AM572510

VERSION AM572510.1 GI:7237243

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 417)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

Tissue Procurement: Christopher Koskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Greg Lennon, Ph.D.
 CDNA Library Arrayed by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml
 Seq primer: -40UP from gibco
 High quality sequence stop: 416.
 Location/Qualifiers

FEATURES

source

```
1. 417
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2751044"
/tissue_type="well-differentiated endometrial
adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/clone_lib="NCI-CGAP-Utl1"
/Note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Life Technologies catalog #:
11538-014"
```

BASE COUNT 140 a 65 c 80 g 132 t

ORIGIN

Query Match 5.1%; Score 105; DB 9; Length 417;
 Best Local Similarity 100.0%; Pred. No. 2.3e-32;
 Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1930 TCCCATCTGTAACTAAATCCCAATTAAGGTCGTTAGAAATGCCCTTTTATG 1989
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 266 TCCCATCTGTAACTAAATCCCAATTAAGGTCGTTAGAAATGCCCTTTTATG 207
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1990 CTCTTAATTATAGCACTAAATGTCATTTTATGAGATCCCAA 2034
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 206 CTCTTAATTATAGCACTAAATGTCATTTTATGAGATCCCAA 162
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 12 358 bp mRNA linear EST 16-OCT-2000
 BF056840/c 7k10d12.x1 NCI-CGAP-CC6 Homo sapiens CDNA clone IMAGE:3443783 3',
 LOCUS mRNA sequence.
 DEFINITION BF056840.1 GI:10810736
 ACCESSION EST.
 VERSION Homo sapiens (human)
 KEYWORDS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 SOURCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 ORGANISM National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 COMMENT Unpublished
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaps-r@mail.nih.gov
 Tissue Procurement: Christopher A. Koskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldi, Ph.D.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
inf@image.llnl.gov
 Seq primer: -40UP from gibco.

FEATURES
 source
 1. 358
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"

BASE COUNT 122 a 78 c 89 g 130 t

/db_xref="taxon:9606"
 /clone="IMAGE:3443783"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /clone_lib="NCI-CGAP-CC6"
 /Note="Vector: p773D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
 from the normalized library NCI-CGAP-CC4 was prepared, and
 ss circles were made in vitro. Following HAP purification,
 this DNA was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from a pool
 of 5,000 clones made from the same library (clones
 1257096-1258631, 1469064-1470983, and 1475592-1476743).
 Subtraction by Bento Soares and M. Fatima Bonaldi."

BASE COUNT 107 a 68 c 82 g 101 t

ORIGIN

Query Match 4.9%; Score 101; DB 10; Length 358;
 Best Local Similarity 100.0%; Pred. No. 1.1e-30;
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1932 CCATCTGTAACTAAATCCCAATTAAGGTCGTTAGAAATGCCCTTTTATGCT 1991
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 101 CCATCTGTAACTAAATCCCAATTAAGGTCGTTAGAAATGCCCTTTTATGCT 42
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1992 TCTTAATTATAGCACTAAATGTCATTTTATGAGATCCT 2032
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 41 TCTTAATTATAGCACTAAATGTCATTTTATGAGATCCT 1

RESULT 13 419 bp mRNA linear EST 13-MAR-2000
 AW572390/c x009a06.x2 NCI-CGAP-Col4 Homo sapiens CDNA clone IMAGE:2799634 3',
 LOCUS mRNA sequence.
 DEFINITION AW572390.1 GI:7237123
 ACCESSION EST.
 VERSION Homo sapiens (human)
 KEYWORDS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 SOURCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 ORGANISM National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 COMMENT Unpublished
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaps-r@mail.nih.gov
 Tissue Procurement: Christopher Koskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml
 Seq primer: -40UP from gibco
 High quality sequence stop: 343.
 Location/Qualifiers

FEATURES
 source
 1. 419
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2799634"
 /tissue_type="moderately-differentiated adenocarcinoma"
 /lab_host="DH10B"
 /clone_lib="NCI-CGAP-Col4"
 /Note="Organ: colon; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.7 kb. Life Technologies catalog #:
 11531-019"

BASE COUNT 122 a 78 c 89 g 130 t

ORIGIN

Query Match 3.6% Score 74; DB 9; Length 419;
 Best Local Similarity 100.0%; Pred. No. 1.4e-19;
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1630 TCCCTGTCACATCTTGCCCAATGGGAGATGATCTTTCACCAAGAGCTCACCAGCAT 1689
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 419 TCCCTGTCACATCTTGCCCAATGGGAGATGATCTTTCACCAAGAGCTCACCAGCAT 360
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1690 TTCACAGAGATGC 1703
 ||||||||||||||||
 Db 359 TTCACAGAGATGC 346

RESULT 14
 LOCUS BF197521/c 525 bp mRNA linear EST 03-NOV-2000
 DEFINITION 7084608.x1 NCI-CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3642903 3'
 similar to TR:Q9UBY3 Q9UBY3 N-ACETYLGLUCOSAMINE
 6-O-SULFOTRANSFERASE.;, mRNA sequence.

ACCESSION BF197521 GI:11086670
 VERSION BF197521
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 525)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moshaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL, send email to:
 info@image.lnl.gov
 High quality sequence stop: 451.

FEATURES

source
 1..525
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3642903"
 /lab_host="DH10B"
 /clone_1lb="NCI-CGAP_Kid11"
 /note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 Plasmid DNA from the normalized library NCI-CGAP_Kid3
 prepared, and ss circles were made in vitro. Following HAP
 purification, this DNA was used as tracer in a subtractive
 hybridization reaction. The driver was PCR-amplified cDNAs
 from a pool of 5,000 clones made from the same library
 (clones 13223/6-13231/1, 1456007-1456775, and
 1500552-1502855). Subtraction by Bento Soares and M.
 Fatima Bonaldo."
 BASE COUNT 90 a 163 c 187 g 85 t

Query Match 2.9% Score 59; DB 10; Length 525;
 Best Local Similarity 100.0%; Pred. No. 2e-13;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCGTCGCGCTCTTACAGCAGTGCTGTCAGAGAGGCGCTTCTTCACTCGAG 682
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 260 AGGCGTCGCGCTCTTACAGCAGTGCTGTCAGAGAGGCGCTTCTTCACTCGAG 202

RESULT 15

AI824100/c 620 bp mRNA linear EST 21-DEC-1999
 LOCUS wj46c01.x1 NCI-CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2405856 3'
 DEFINITION similar to TR:075667 075667 DJ7LL16.4; contains PIR5.b2 PIR5
 repetitive element.;, mRNA sequence.

ACCESSION AI824100 GI:5444771
 VERSION AI824100
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 620)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moshaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www-bio.lnl.gov/dbp/image/image.html
 Insert Length: 917 Std Error: 0.00
 Seq primer: -400P from Glibco
 High quality sequence stop: 490.

FEATURES

source
 1..620
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2405856"
 /tissue_type="squamous cell carcinoma, poorly
 differentiated (4 pooled tumors, including primary and
 metastatic)"
 /dev_stage="adult"
 /lab_host="DH10B (phage-resistant)"
 /clone_1lb="NCI-CGAP_Lu19"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 pooled lung tumor tissue, and was then primed with a Not I
 - oligo(dT) primer. Double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. Library went through one round of
 normalization. Library constructed by Bento Soares and M.
 Fatima Bonaldo."
 BASE COUNT 106 a 195 c 227 g 89 t 3 others

Query Match 2.9% Score 59; DB 9; Length 620;
 Best Local Similarity 100.0%; Pred. No. 1.9e-13;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCGTCGCGCTCTTACAGCAGTGCTGTCAGAGAGGCGCTTCTTCACTCGAG 682
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 314 AGGCGTCGCGCTCTTACAGCAGTGCTGTCAGAGAGGCGCTTCTTCACTCGAG 256

Search completed: August 14, 2003, 21:16:41
 Job time : 3900 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 11:38:30 ; Search time 541 Seconds

(Without alignments)
10193.996 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043

Sequence: 1 gaattccattgtgttgta.....tgggattcctaaataaaaaa 2043

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5098080

Minimum DB seq length: 10

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_19Jun03:*

1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*

2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*

3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*

4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*

5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*

6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*

7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*

8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*

9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*

10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*

11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*

12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*

13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*

14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*

15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*

16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*

17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*

18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*

19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*

20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*

21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*

22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*

23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*

24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

25: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1547	75.7	1979	22	AAK94229	Human full-length
2	1547	75.7	2032	20	AAZ20792	Human glycosyl sul
3	1262	61.8	1333	24	AA516947	Human L-selectin s
4	1153	56.4	2065	21	AAZ94211	Human transferrase
5	695	34.0	877	22	AAK91803	Human cDNA 5'-end
6	695	34.0	877	22	AAK93921	Human cDNA clone I
7	403	19.7	505	24	ABV89280	Human colon cancer
8	403	19.7	517	24	ABK54724	Human colon cancer

9	306	15.0	2988	21	AAK76156	Human OREF ORF1711
10	114	5.6	389	22	ABK54794	Human colon cancer
11	105	5.1	548	22	AAK92588	Human cDNA 3'-end
12	59	2.9	1647	24	AAK24670	Human drug metabol
13	59	2.9	1694	22	AAK02700	Human glycosyl sul
14	59	2.9	2044	22	AAK02699	Human glycosyl sul
15	59	2.9	2170	22	AAK02698	Human glycosyl sul
16	59	2.9	2544	24	ABN89505	Human corneal N-ac
17	59	2.9	48435	24	ABN89533	Human corneal N-ac
18	59	2.9	160552	22	AAK02697	Human glycosyl sul
19	37	1.8	37	20	AAZ20798	PCR primer for gly
20	30	1.5	30	24	ABK70821	Probe for human ge
21	27	1.3	1926	20	AAZ20793	Mouse glycosyl sul
22	23	1.1	1937	22	AAK16948	Murine intestinal
23	23	1.1	1989	22	AAK02696	Mouse glycosyl sul
24	22	1.1	22	24	ABK70834	PCR primer for hum
25	22	1.1	22	24	AAK16960	Human L-selectin s
26	22	1.1	22	24	AAK16962	Human L-selectin s
27	22	1.1	22	24	AAK16964	Human L-selectin s
28	22	1.1	167	24	ABL86480	Human ovarian can
29	22	1.1	171	24	ABL86089	Human ovarian can
30	22	1.1	187	24	ABL86207	Human ovarian can
31	22	1.1	191	24	ABL86306	Human ovarian can
32	21	1.0	21	24	ABK70835	PCR primer for hum
33	21	1.0	21	24	AAK16961	Human L-selectin s
34	21	1.0	10004	22	ABK14483	Human nervous syst
35	21	1.0	10246	22	ABK14485	Human nervous syst
36	20	1.0	88	21	AAK25980	Human secreted pro
37	20	1.0	354	22	AAK55813	Novel human polynu
38	20	1.0	370	22	AAK55341	Novel human polynu
39	20	1.0	381	22	AAK67389	Novel human polynu
40	20	1.0	398	25	AAK45934	Bovine EST associa
41	20	1.0	3763	19	AAK58194	Human myosin I-cha
42	20	1.0	3763	23	AAK59872	DNA encoding novel
43	20	1.0	5814	23	AAK87855	Human ovary specif
44	20	1.0	6147	25	AAK64533	Human dlthp protei
45	20	1.0	9516	22	AAK73555	Human immune/haema

ALIGNMENTS

RESULT 1	AAK94229	standard; cDNA: 1979 BP.
AC	AAK94229	
XX		
DT	06-NOV-2001	(first entry)
XX		
DE	Human full-length cDNA, SEQ ID NO: 2816.	
XX		
KW	Human; full length cDNA; cDNA synthesis; oligo-capping; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	EP1130094-A2.	
XX		
PD	05-SEP-2001.	
XX		
PF	07-JUL-2000; 2000EP-0114089.	
XX		
PR	08-JUL-1999; 99JP-0194486.	
XX		
PR	11-JAN-2000; 2000JP-0118774.	
PR	02-MAY-2000; 2000JP-0183765.	
XX		
PA	(HELI-) HELIX RES INST.	
XX		
PI	Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;	
XX	Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;	
XX	WPI: 2001-524255/58.	
DR	P-PSDB; AAM93309.	

XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8: SEQ ID NO 2816; 1380bp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a full length
CC human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 1979 BP; 451 A; 542 C; 474 G; 512 T; 0 other;

Query Match 75.7%; Score 1547; DB 22; Length 1979;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 106 CAGGCTCTCCACTTCAGACAAATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTCT 165
DB 47 CAGGCTCTCCACTTCAGACAAATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTCT 106
QY 166 GGTTCCTCCAGATGGCATCTTGCTCTATCTTCCACATGTACAGCCACAAATCAGCTC 225
DB 107 GGTTCCTCCAGATGGCATCTTGCTCTATCTTCCACATGTACAGCCACAAATCAGCTC 166
QY 226 CCGTCTATGAAGGACAGCCCGAGGAGCATGACGTGCTGCTCTCTCTGCGCTC 285
DB 167 CCGTCTATGAAGGACAGCCCGAGGAGCATGACGTGCTGCTCTCTCTGCGCTC 226
QY 286 TGGCTCTCTTTTGGGGGACCTTTTGGGACAGCCAGATGTTTCTTACCTGATGA 345
DB 227 TGGCTCTCTTTTGGGGGACCTTTTGGGACAGCCAGATGTTTCTTACCTGATGA 286
QY 346 GCCCGCTGGACAGTGTGATGACCTTCAAGCAGACACCGCTGGATCTGACATGGC 405
DB 287 GCCCGCTGGACAGTGTGATGACCTTCAAGCAGACACCGCTGGATCTGACATGGC 346
QY 406 TGTGGGGGATGTGAACGGGCGCTCTTCTTGGCAGCATGAGCGTCTTGGATGCTACAT 465
DB 347 TGTGGGGGATGTGAACGGGCGCTCTTCTTGGCAGCATGAGCGTCTTGGATGCTACAT 406
QY 466 GGAACCTGCTCCCCGGAGACATCCAGCTCTTTCAGTGGGAGAAAGCGGCGCTGTG 525
DB 407 GGAACCTGCTCCCCGGAGACATCCAGCTCTTTCAGTGGGAGAAAGCGGCGCTGTG 466
QY 526 TTTCTACACCTGCTGTGACATCATCCACAAATGAAATCATCCCGGGGCTCAGTCAG 585
DB 467 TTTCTACACCTGCTGTGACATCATCCACAAATGAAATCATCCCGGGGCTCAGTCAG 526
QY 586 GCTCTGTGACATCAACAGCCCTTGAAGTGTGAGAAAGGCGCTGCGCTCTACAGCA 645
DB 527 GCTCTGTGACATCAACAGCCCTTGAAGTGTGAGAAAGGCGCTGCGCTCTACAGCA 586
QY 646 CTTGCTGCTCAAGGAGGTGCGCTTCTCAACTGCAGTCCCTCTACCCGCTGTGAAGA 705
DB 587 CTTGCTGCTCAAGGAGGTGCGCTTCTCAACTGCAGTCCCTCTACCCGCTGTGAAGA 646
QY 706 CCCCCTCCCTCAACCTGCATATGCTGACCTGCTCCGGAGCCCGGGGCGCTGTGCTTC 765
DB 647 CCCCCTCCCTCAACCTGCATATGCTGACCTGCTCCGGAGCCCGGGGCGCTGTGCTTC 706
QY 766 CCGAAGACGACAAAGGAGATCTCATGTTGACAGTCCATTTGATGGGCGAGCATGA 825
DB 707 CCGAAGACGACAAAGGAGATCTCATGTTGACAGTCCATTTGATGGGCGAGCATGA 766

QY 826 GCATAAACTCAGAGAGAGACCAACCTACTATGTGATGACAGTCACTGCCAAAGCA 885
DB 767 GCATAAACTCAGAGAGAGACCAACCTACTATGTGATGACAGTCACTGCCAAAGCA 826
QY 886 GCTGGAGATCTCAAAACCATCCAGTCTTGGCCCAAGGCGCTGACAGAAAGCTACTGCT 945
DB 827 GCTGGAGATCTCAAAACCATCCAGTCTTGGCCCAAGGCGCTGACAGAAAGCTACTGCT 886
QY 946 TGTGCGCTTGAAGAGACCTGCTGAGCCCTGTGCGCCAGACTTCCGAAATGATGAAT 1005
DB 887 TGTGCGCTTGAAGAGACCTGCTGAGCCCTGTGCGCCAGACTTCCGAAATGATGAAT 946
QY 1006 CCGGGATTTGGAATTTTGGCCCATTTTGAAGACTGGGTCATACATCAGCCAGAGCA 1065
DB 947 CCGGGATTTGGAATTTTGGCCCATTTTGAAGACTGGGTCATACATCAGCCAGAGCA 1006
QY 1066 GGGCATGGGTGACACAGCTTTTCCACAAATGACAGAGATGCCCTTAATGTGCCAGGC 1125
DB 1007 GGGCATGGGTGACACAGCTTTTCCACAAATGACAGAGATGCCCTTAATGTGCCAGGC 1066
QY 1126 TTGGCGCTGCTTTTGGCCCTATGAAAAAGTTTCTGACCTTCAGAAAAAGCTGTGCGATGC 1185
DB 1067 TTGGCGCTGCTTTTGGCCCTATGAAAAAGTTTCTGACCTTCAGAAAAAGCTGTGCGATGC 1126
QY 1186 CATGAATTTGCTGGCTACCCGCCACGTCAGATGTGAACAAGAACAGAAACCTGTTGCT 1245
DB 1127 CATGAATTTGCTGGCTACCCGCCACGTCAGATGTGAACAAGAACAGAAACCTGTTGCT 1186
QY 1246 GGAATCTTCTGCTACCTGACCTGCTCCTAGCAAGAAATCCACTAAGAGGTTGAGAAAGCTT 1305
DB 1187 GGAATCTTCTGCTACCTGACCTGCTCCTAGCAAGAAATCCACTAAGAGGTTGAGAAAGCTT 1246
QY 1306 TGTCTGACACCTGCTGCTACACCTTCAGTCACTTCTCTGAATGCTTGTAGCCCTGTGCTAC 1365
DB 1247 TGTCTGACACCTGCTGCTACACCTTCAGTCACTTCTCTGAATGCTTGTAGCCCTGTGCTAC 1306
QY 1366 TCTCTAGGCTTAACATCACTGTCTGCTGATACACTAGTGTGATGCTGTGCTCAGCAG 1425
DB 1307 TCTCTAGGCTTAACATCACTGTCTGCTGATACACTAGTGTGATGCTGTGCTCAGCAG 1366
QY 1426 TGTCTAGGAGAGAAAGCTTTTGTCTCAGTCTGTGTAGAAACAGACTGGGAAAGCT 1485
DB 1367 TGTCTAGGAGAGAAAGCTTTTGTCTCAGTCTGTGTAGAAACAGACTGGGAAAGCT 1426
QY 1486 TATGTAGAGACACATCCACAGTGAAGAGGATGCTCTTCTTCTTCTTCTGATCT 1545
DB 1427 TATGTAGAGACACATCCACAGTGAAGAGGATGCTCTTCTTCTTCTTCTGATCT 1486
QY 1546 TCTCTGTGCGGAGACTTCAAGAGACTTGTGGCTGAGAGCCATTAACAGACAGCAGT 1605
DB 1487 TCTCTGTGCGGAGACTTCAAGAGACTTGTGGCTGAGAGCCATTAACAGACAGCAGT 1546
QY 1606 ATCAGTGAATGATTCATTAACCTCCCTGTCACACTTGTGCCAATGGGAGATGATCT 1665
DB 1547 ATCAGTGAATGATTCATTAACCTCCCTGTCACACTTGTGCCAATGGGAGATGATCT 1606
QY 1666 TTACCAAAAGAGCTACAGCACTTTTCCACAGAGATGC 1703
DB 1607 TTACCAAAAGAGCTACAGCACTTTTCCACAGAGATGC 1644

RESULT 2
AA220792
ID AA220792 standard; DNA; 2032 BP.
XX
AC AA220792;
XX
DT 08-DEC-1999 (first entry)
XX
DE Human glycosyl sulfotransferase-3 coding sequence.
XX
KW glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
selectin binding interaction; inflammation; lymphocyte homing; human;

secondary lymph organ; ss.
Homo sapiens.
W09949018-A1.
30-SEP-1999.
26-FEB-1999; 99MO-US04316.
20-MAR-1998; 98US-0045284.
12-NOV-1998; 98US-0190911.
(REGC) UNIV CALIFORNIA.
(SYNT) SYNTX USA INC.
Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
WPI; 1999-580442/49.
P-PSDB; AAY39918.
Human and murine glycosyl sulfotransferase 3 and related polynucleotides
Claim 4; Fig 1; 59pp; English.
This sequence encodes the human glycosyl sulfotransferase-3 (GST-3) of the invention. The nucleic acid sequences, probes and primers derived from these, proteins and antibodies are useful in detecting homologues. The sequences, antibodies and methods are useful in the diagnosis and treatment of diseases associated with selectin binding interactions, including conditions associated with or resulting from the homing of leukocytes to sites of inflammation and the normal homing of lymphocytes to secondary lymph organs.
Sequence 2032 BP; 468 A; 569 C; 490 G; 505 T; 0 other.
Query Match 75.7%; Score 1547; DB 20; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 106 CAAAGCTTCCACCTTCAAGCAGCATGCTACCTGCTTAAAGCTCTGCTGTTTCT 165
DB 160 CAAAGCTTCCACCTTCAAGCAGCATGCTACCTGCTTAAAGCTCTGCTGTTTCT 219
QY 166 GGTTCACAGATGAGCAGCTTGTGCTATCTTCCACATGACGACCAACATCAGCTC 225
DB 220 GGTTCACAGATGAGCAGCTTGTGCTATCTTCCACATGACGACCAACATCAGCTC 279
QY 226 CTTGCTATGAAAGCAGACGCGGAGCGATGACGCTGCTGCTTCTGCTGCTGCTC 285
DB 280 CTTGCTATGAAAGCAGACGCGGAGCGATGACGCTGCTGCTGCTGCTGCTGCTC 339
QY 286 TGGCTCTTCTTGTGGGGAGCTTTTGGGAGCAGCACCAGATGTTTCTACGTGAGA 345
DB 340 TGGCTCTTCTTGTGGGGAGCTTTTGGGAGCAGCACCAGATGTTTCTACGTGAGA 399
QY 346 GCCCGCTGGCAGCTGGATGACCTTCAAGCAGACGCGCTGGATGCTGACATGCGC 405
DB 400 GCCCGCTGGCAGCTGGATGACCTTCAAGCAGACGCGCTGGATGCTGACATGCGC 459
QY 406 TGTGCGGATCTGATAGGCGGCTTCTTGTGACATGAGCTCTTGTGATGCTACAT 465
DB 460 TGTGCGGATCTGATAGGCGGCTTCTTGTGACATGAGCTCTTGTGATGCTACAT 519
QY 466 GAAACCTGTGCTCCGGAGACAGTCCAGCTCTTTCAGTGGAGAAACAGCGGCGCTGTG 525
DB 520 GAAACCTGTGCTCCGGAGACAGTCCAGCTCTTTCAGTGGAGAAACAGCGGCGCTGTG 579
QY 526 TTTCGACCTGCTGCTGATCATCCACAAGATGAATCATCCCGGGGCTCTACGTAG 585
DB 580 TTTCGACCTGCTGCTGATCATCCACAAGATGAATCATCCCGGGGCTCTACGTAG 639

QY 586 GCTCTGTGACATCAAGCCCTTTGAGGTGGTGGAGAGGCTGCGCTCTCTACAGCA 645
DB 640 GCTCTGTGACATCAAGCCCTTTGAGGTGGTGGAGAGGCTGCGCTCTCTACAGCA 699
QY 646 CGTGTGCTCAAGAGGTGCGCTTCTTCAACCTCAGTCCCTCAACCGCTGTGAAGA 705
DB 700 CGTGTGCTCAAGAGGTGCGCTTCTTCAACCTCAGTCCCTCAACCGCTGTGAAGA 759
QY 706 CCCCTCCCTCAACCTCAGTATGCTGACCTGCTCCGGAGACCCCGGGCGCTTCCGCTC 765
DB 760 CCCCTCCCTCAACCTCAGTATGCTGACCTGCTCCGGAGACCCCGGGCGCTTCCGCTC 819
QY 766 CCGAGAACGACCAAGAGGAGATCTCATGATGACAGTGGCATTTGATGGAGGACATGA 825
DB 820 CCGAGAACGACCAAGAGGAGATCTCATGATGACAGTGGCATTTGATGGAGGACATGA 879
QY 826 GCAAAAACCTCAAGAGAGAGACCAACCTTATGTGATGAGTCAATCTCCCAAGCCA 885
DB 880 GCAAAAACCTCAAGAGAGAGACCAACCTTATGTGATGAGTCAATCTCCCAAGCCA 939
QY 886 GCTGAGATCTCAAGAGACCATCCAGTCTCTGCGGACCCCTGAGAGAGCTACCTGCT 945
DB 940 GCTGAGATCTCAAGAGACCATCCAGTCTCTGCGGACCCCTGAGAGAGCTACCTGCT 999
QY 946 TGTGCGCTATGAGACCTGCTGAGCCCTGTGGCCAGACTTCCGGAATGATGAAT 1005
DB 1000 TGTGCGCTATGAGACCTGCTGAGCCCTGTGGCCAGACTTCCGGAATGATGAAT 1059
QY 1006 CGTGGATTTGAATTTCTTGGCCCATCTTCAAGACCTGGGTGATTAACATCCAGGCA 1065
DB 1060 CGTGGATTTGAATTTCTTGGCCCATCTTCAAGACCTGGGTGATTAACATCCAGGCA 1119
QY 1066 GGGCATGGGTACCAAGCTTTCACACAAAGGAGGATCCCTTATGTCTCCAGGC 1125
DB 1120 GGGCATGGGTACCAAGCTTTCACACAAAGGAGGATCCCTTATGTCTCCAGGC 1179
QY 1126 TTGGCGGTGCTTGTGCTTGAAGAAAGTTCTGATGAGAAAGCTGTGGGATGC 1185
DB 1180 TTGGCGGTGCTTGTGCTTGAAGAAAGTTCTGATGAGAAAGCTGTGGGATGC 1239
QY 1186 CATGAATTTGCTGGGCTACCGCCACATCTGATCTGACAAAGACAGAAACCTGTCT 1245
DB 1240 CATGAATTTGCTGGGCTACCGCCACATCTGATCTGACAAAGACAGAAACCTGTCT 1299
QY 1246 GGATCTTCTGCTACCTGAGTGCCTGAGCAATCAGTAAGGGTTGAGAAAGCTT 1305
DB 1300 GGATCTTCTGCTACCTGAGTGCCTGAGCAATCAGTAAGGGTTGAGAAAGCTT 1359
QY 1306 TGTGCGACCTGCTGCTAGCTCACTTCTCTGAATGCTTCTGAGCCTTGGCTACA 1365
DB 1360 TGTGCGACCTGCTGCTAGCTCACTTCTCTGAATGCTTCTGAGCCTTGGCTACA 1419
QY 1366 TCTCTGAGCCTTAACTACATCTGTGGGTATCACACTGATGTGATGTGTCCACAG 1425
DB 1420 TCTCTGAGCCTTAACTACATCTGTGGGTATCACACTGATGTGATGTGTGTCCACAG 1479
QY 1426 TGTCAAGCAGAAAGACTTGTGTGCTAGTGTGTGTAAGAAACAGACTGGGAACT 1485
DB 1480 TGTCAAGCAGAAAGACTTGTGTGCTAGTGTGTGTAAGAAACAGACTGGGAACT 1539
QY 1486 TATGTAGACACACATCCACAGTGAAGAGGATATGCTCTTCTTCTTCTTGATCT 1545
DB 1540 TATGTAGACACACATCCACAGTGAAGAGGATATGCTCTTCTTCTTCTTGATCT 1599
QY 1546 TCTGTCTGGGACAGACTTCAAGACTTGTGGCTGGAGGCTTATTAAGCAGACAGT 1605
DB 1600 TCTGTCTGGGACAGACTTCAAGACTTGTGGCTGGAGGCTTATTAAGCAGACAGT 1659
QY 1606 ATCAGTGAATTTGATCAATAACCTCTGTCACATCTGCCAAATGGGAAAGGATCT 1665
DB 1660 ATCAGTGAATTTGATCAATAACCTCTGTCACATCTGCCAAATGGGAAAGGATCT 1719
QY 1666 TTCAACAAAGAGCTACACAGCATTTTCCACAGAGATGC 1703

DB 1720 TTCACCAAGAGCTCAGCAGCATTTTCACAGAGATGC 1757

RESULT 3

AA16947
ID AA16947 standard; cDNA; 1333 BP.

AA16947;

12-MAR-2002 (first entry)

Human L-selectin sulfotransferase-2 (LSST-2) cDNA.

Human; beta1,3gnt; beta1,3-N-acetylglucosaminyltransferase; MECA-79; ss;
L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;
allergic contact dermatitis; lymphoma; chronic pneumonia; LSST-2;
delayed-type hypersensitivity reaction; hyperplastic thymus; antileuker;
antiinflammatory; antipsoriatic; antidiabetic; dermatological;
antiallergic.

Homo sapiens.

Key Location/Qualifiers
CDS 111..1253
/tag a
/product "Human LSST-2"

MO200185177-A1.

15-NOV-2001.

10-MAY-2001; 2001WO-US15452.

11-MAY-2000; 2000US-0569320.

(BURN-) BURNHAM INST.

Fukuda M, Yeh J, Hiraoaka N;

WPI: 2002-075226/10.

P-PSDB; AAU11274.

New enzyme, useful for modifying acceptor molecule, comprises an
isolated L-selectin sulfotransferase-2 that directs expression of
L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
intestinal GLCNAc 6-sulfotransferase

Claim 19; Fig 4; 98pp; English.

The present invention provides a method of modifying an acceptor molecule
by contacting the acceptor with an isolated

fragment, where beta1,3gnt directs expression of a MECA-79 antigen. The
invention also provides a method of treating or preventing an
L-selectin-mediated condition by reducing the expression or activity of a
beta1,3gnt that directs expression of a MECA-79 antigen. This can be done
by administering to the subject an oligosaccharide L-selectin antagonist
that inhibits binding of L-selectin to a MECA-79 antigen, for example by
administering antibody material that specifically binds beta1,3gnt,
and/or a beta1,3gnt antisense nucleic acid molecule. L-selectin
sulfotransferase-2 (LSST-2) also directs MECA-79 antigen expression.
Alternatively, the expression or activity of LSST-2 or its active
fragment can be reduced in combination with reducing the expression or
activity of beta1,3gnt. The method is useful for treating L-selectin
mediated conditions such as Crohn's disease and ulcerative colitis,
inflammatory disorders of the skin such as allergic contact dermatitis,
psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
hypersensitivity reactions, diabetes and hyperplastic thymus. This
sequence represents cDNA encoding human LSST-2.

Sequence 1333 BP; 286 A; 393 C; 336 G; 318 T; 0 other;

Query Match 61.8%; Score 1262; DB 24; Length 1333;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	107	AAAGTTTCCACTTCAGCACAATGCTACTGCTCTTAAAAAATGAAGTCTGCTGTTCTG	166
DB	72	AAAGTTTCCACTTCAGCACAATGCTACTGCTCTTAAAAAATGAAGTCTGCTGTTCTG	131
QY	167	GTTTCCAGATGGCCCTTGGCTATTTCTTCCACATGTACAGCCACAACATCAAGCTCC	226
DB	132	GTTTCCAGATGGCCCTTGGCTATTTCTTCCACATGTACAGCCACAACATCAAGCTCC	191
QY	227	CTGTCTATGAAGGACAGCCGAGCCATGACAGTGTCTGTCTTCTGAGGCTCT	286
DB	192	CTGTCTATGAAGGACAGCCGAGCCATGACAGTGTCTGTCTTCTGAGGCTCT	251
QY	287	GGCTCTTCTTCTTGTGGGAGCTTTTGGGACAGCCACAGATGTTTCTACCTGATGAG	346
DB	252	GGCTCTTCTTCTTGTGGGAGCTTTTGGGACAGCCACAGATGTTTCTACCTGATGAG	311
QY	347	CCGCGCTGGCAGTGTGATGACCTTCAAGCAGACGCGCGATGCTGACATAGGCT	406
DB	312	CCGCGCTGGCAGTGTGATGACCTTCAAGCAGACGCGCGATGCTGACATAGGCT	371
QY	407	GTCGGGATCTGATACGGGCGCTTCTTGTGCGATGAGCGCTTTGATGCTACATG	466
DB	372	GTCGGGATCTGATACGGGCGCTTCTTGTGCGATGAGCGCTTTGATGCTACATG	431
QY	467	GAACCTGTGTCGGGAGAGATGACAGCTCTTCACTGAGGAGAAACAGCCGCGCTGT	526
DB	432	GAACCTGTGTCGGGAGAGATGACAGCTCTTCACTGAGGAGAAACAGCCGCGCTGT	491
QY	527	TCTGACCTGCTGTGATGATCCCAAGATGAAATATGATCCCGGCGCTCACTGACAG	586
DB	492	TCTGACCTGCTGTGATGATCCCAAGATGAAATATGATCCCGGCGCTCACTGACAG	551
QY	587	CTCCTGTGACATCAACAGCCCTTGTGAGTGTGAGAAAGGCTGCCCTCTACAGCCAC	646
DB	552	CTCCTGTGACATCAACAGCCCTTGTGAGTGTGAGAAAGGCTGCCCTCTACAGCCAC	611
QY	647	GTCGTGCTCAAGAGTGTGCTTCTCAACCTGAGTGTGAGAAAGGCTGCCCTCTACAG	706
DB	612	GTCGTGCTCAAGAGTGTGCTTCTCAACCTGAGTGTGAGAAAGGCTGCCCTCTACAG	671
QY	707	CCCTCCCTCAACCTGATATGCTGACCTGTGCGGAGACCCCGGCGCTGTCCGTTCC	766
DB	672	CCCTCCCTCAACCTGATATGCTGACCTGTGCGGAGACCCCGGCGCTGTCCGTTCC	731
QY	767	CGAGAAAGCACAAGGAGATCTCATGATTGACAGTGTGATGAGGAGCAGATGAG	826
DB	732	CGAGAAAGCACAAGGAGATCTCATGATTGACAGTGTGATGAGGAGCAGATGAG	791
QY	827	CAAAAACCTAAGAGAGAGACCAACCTCTATGATGATGATGATGATGATGATGATG	886
DB	792	CAAAAACCTAAGAGAGAGACCAACCTCTATGATGATGATGATGATGATGATGATG	851
QY	887	CTGAGATCTACAGACATCCAGTCTTGTCCCAAGGCGCTCGAGAAAGCTACCTGCTT	946
DB	852	CTGAGATCTACAGACATCCAGTCTTGTCCCAAGGCGCTCGAGAAAGCTACCTGCTT	911
QY	947	GTCGCTATGAGAGCTGTGCTGTGAGCCCTGTGAGCCCAAGCTTCCGAATGATGATTC	1006
DB	912	GTCGCTATGAGAGCTGTGCTGTGAGCCCTGTGAGCCCAAGCTTCCGAATGATGATTC	971
QY	1007	GTCGCTATGAGAGCTGTGCTGTGAGCCCTGTGAGCCCAAGCTTCCGAATGATGATTC	1066
DB	972	GTCGCTATGAGAGCTGTGCTGTGAGCCCTGTGAGCCCAAGCTTCCGAATGATGATTC	1031
QY	1067	GGCATGGGTGACAGCTTTCACACAATGCGAGGATGCGCTTAATGTCTCCAGGCT	1126
DB	1032	GGCATGGGTGACAGCTTTCACACAATGCGAGGATGCGCTTAATGTCTCCAGGCT	1091

QY 1127 TGGGCTGCTCTTGGCCCTATGAAAGGTTCTGACTTCAGACGAAAGCCTTGCGGATGCC 1186
 |||||||
 DB 1092 TGGGCTGCTCTTGGCCCTATGAAAGGTTCTGACTTCAGACGAAAGCCTTGCGGATGCC 1151
 |||||||
 QY 1187 ATGAATTTGCTGGGCTACCCGACAGTCTGATCTGAACAAGAACAAACCTGTGCTG 1246
 |||||||
 DB 1152 ATGAATTTGCTGGGCTACCCGACAGTCTGATCTGAACAAGAACAAACCTGTGCTG 1211
 |||||||
 QY 1247 GATCTTCTGCTACCTGAGTCTGCTGAGCAATTCACCTAAGAGGGTTGAAGGCTTT 1306
 |||||||
 DB 1212 GATCTTCTGCTACCTGAGTCTGCTGAGCAATTCACCTAAGAGGGTTGAAGGCTTT 1271
 |||||||
 QY 1307 GCTGCCACCTGGTGTACAGCTCAGTCACTTCTCTGAATGCTTCTGAGCCTTGCTACAT 1366
 |||||||
 DB 1272 GCTGCCACCTGGTGTACAGCTCAGTCACTTCTCTGAATGCTTCTGAGCCTTGCTACAT 1331
 |||||||
 QY 1367 CT 1368
 ||
 DB 1332 CT 1333

RESULT 4

AA294211 standard; cDNA: 2065 BP.

AA294211:

19-JUN-2000 (first entry)

Human transferase TRNSFS-11 cDNA clone 2617407CBL.

Transferase; TRNSFS-11; human; antitumor; cell proliferation;
 inflammation; gastrointestinal disorder; developmental disorder;
 genetic disorder; neurological disorder; reproductive disorder;
 smooth muscle disorder; immunological disorder; gene therapy;
 diagnosis; N-acetylglycosamine 6-O-sulfotransferase; ss.

Homo sapiens.

Location/Qualifiers

Key 174..1334

FT CDS /tag= a

PN WO200014251-A2.

PD 16-MAR-2000.

PF 09-SEP-1999; 99WO-US20989.

PR 10-SEP-1998; 98US-0150657.

PR 04-NOV-1998; 98US-0186779.

PR 11-MAY-1999; 99US-0133642.

PA (INCY-) INCYTE PHARM INC.

PI Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;

PI Hillman JL, Azimzal Y;

DR WPI: 2000-256996/22.

DR P-PSDB; AAY79219.

XX Human transferase proteins useful for preventing, diagnosing and

XX treating cancers and developmental, gastrointestinal, genetic,

XX immunological, neurological, reproductive and smooth muscle disorders -

XX Claim 9; Page 104-105; 113pp; English.

XX The present sequence is that of cDNA clone 2617407CBL encoding

XX human transferase TRNSFS-11 (see AAY79219), 1 of 15 claimed human

XX transferase proteins of the invention (see AAY79209-23). The clone

XX was isolated from gall bladder cDNA library GBLANOT01. TRNSFS-11

XX is expressed in dermatologic and gastrointestinal tissues,

XX especially those associated with inflammation and cell

CC proliferation. It shows homology to mouse N-acetylglycosamine
 CC 6-O-sulfotransferase. A polynucleotide comprising nucleotides
 CC 264-332 or 1272-1331 of the present sequence can be used as a DNA
 CC probe. The new human transferases and polynucleotides can be used
 CC in the diagnosis, prevention and treatment (including gene therapy
 CC and antisense therapy) of cancer, developmental disorders,
 CC gastrointestinal disorders, genetic disorders, immunological
 CC disorders, neurological disorders, reproductive disorders, and
 CC smooth muscle disorders.

SQ Sequence 2065 BP; 466 A; 575 C; 491 G; 533 T; 0 other;

Query Match 56.4%; Score 1153; DB 21; Length 2065;

Best Local Similarity 99.6%; Pred. NO. 0;

Matches 1593; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 106 CAAGCTCTTCCACTTCACACACATGCTCTGCTTAAAAAATGAAGCTCTGCTTCT 165
 |||||||
 DB 152 CAAGCTCTTCCACTTCACACACATGCTCTGCTTAAAAAATGAAGCTCTGCTTCT 211
 |||||||
 QY 166 GGTTCGCCAGATGGCCATCTTGGCTCTATCTTCACATGTACAGCCCAACATCAGCTC 225
 |||||||
 DB 212 GCTTCCAGATGGCCATCTTGGCTCTATCTTCACATGTACAGCCCAACATCAGCTC 271
 |||||||
 QY 226 CCTGTCTATGAAGCACAGCCGACGATGCTGCTGCTCTCTCTGCTGCTGCTGCTGCT 285
 |||||||
 DB 272 CCTGTCTATGAAGCACAGCCGACGATGCTGCTGCTCTCTCTGCTGCTGCTGCTGCTGCT 331
 |||||||
 QY 286 TGGCTCTTCTTTTGGGGGACGCTTTTGGGACAGCCAGATGCTTTCTTCTACCTATGGA 345
 |||||||
 DB 332 TGGCTCTTCTTTTGGGGGACGCTTTTGGGACAGCCAGATGCTTTCTTCTACCTATGGA 391
 |||||||
 QY 346 GCCGCCGCGGACGCTGTGATGACCTTCAAGACAGACGCGCGGTGATGCTGACATGAC 405
 |||||||
 DB 392 GCCGCCGCGGACGCTGTGATGACCTTCAAGACAGACGCGCGGTGATGCTGACATGAC 451
 |||||||
 QY 406 TGTGGGATGTGATACGCGGCTCTTCTTGTGACATGACGCTTTTGTATGCTTACAT 465
 |||||||
 DB 452 TGTGGGATGTGATACGCGGCTCTTCTTGTGACATGACGCTTTTGTATGCTTACAT 511
 |||||||
 QY 466 GGAACCTGTGCTGCGGACAGTCCAGCTCTTTTGTAGTGGAGAACAGCCGGGCTGTG 525
 |||||||
 DB 512 GGAACCTGTGCTGCGGACAGTCCAGCTCTTTTGTAGTGGAGAACAGCCGGGCTGTG 571
 |||||||
 QY 526 TTTGTCACCTGCTGTGATCATATCCCAAGATGAATCCCGCGGGCTCACTGAG 585
 |||||||
 DB 572 TTTGTCACCTGCTGTGATCATATCCCAAGATGAATCCCGCGGGCTCACTGAG 630
 |||||||
 QY 586 GCTCCTGTGACATCAAGCCCTTTG-AGTGTGTGAGAGAGCCCTGCTCTACAGCC 644
 |||||||
 DB 631 GCTCCTGTGACATCAAGCCCTTTGAGTGTGTGAGAGAGCCCTGCTCTACAGCC 690
 |||||||
 QY 645 AGTGTGTGACATCAAGAGGTGCGCTTTCAACCTGACAGTCCCTTACCCGCTGTAAG 704
 |||||||
 DB 691 AGTGTGTGACATCAAGAGGTGCGCTTTCAACCTGACAGTCCCTTACCCGCTGTAAG 750
 |||||||
 QY 705 ACCCTGCTCAACCTGATATCGACACCTGCTGCGGAGACCCCGGGCCGCTGCTT 764
 |||||||
 DB 751 ACCCTGCTCAACCTGATATCGACACCTGCTGCGGAGACCCCGGGCCGCTGCTT 810
 |||||||
 QY 765 CCCGAGACGCAAGAGGAGATCTCATGATTTGACATGCGATTTGATGGGCGAGCATG 824
 |||||||
 DB 811 CCCGAGACGCAAGAGGAGATCTCATGATTTGACATGCGATTTGATGGGCGAGCATG 870
 |||||||
 QY 825 AGCAAAACTCAAGAGAGGAGACCAACCTCATATGATGATGAGGTATCTGCAAAACC 884
 |||||||
 DB 871 AGCAAAACTCAAGAGAGGAGACCAACCTCATATGATGATGAGGTATCTGCAAAACC 930
 |||||||
 QY 885 AGCTGAGATCTTACAGACATCCAGTCTTCCCAAGGCGCTGAGGAGATGATCCGCG 944
 |||||||
 DB 931 AGCTGAGATCTTACAGACATCCAGTCTTCCCAAGGCGCTGAGGAGATGATCCGCG 990
 |||||||
 QY 945 TTGTGCGCTATGAGACCTGCTGAGCCCTGTGTCAGCACTTCCGGAATGTATGAAT 1004

```

Db      991 TTGTGGCTATAGAGGACCTGGCTCGAGCCCTGTGGCCAGACTTCCCAATGTATGAAT
OY      1005 TCGTGGGATTTGGAAATCTTGGCCCATCTTCAGACCTGGGATGACATACATCACCCGGGCA
Db      1051 TCGTGGGATTTGGAAATCTTGGCCCATCTTCAGACCTGGGATGACATACATCACCCGGGCA
OY      1065 AAGGCAATGGGTAGCAGCAGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGG
Db      1111 AAGGCAATGGGTAGCAGCAGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGG
OY      1125 CTGGGGCTGGTCTTGGCCCTATGAAAAGTTCTTGACTTCAGAAAAGCTGTGGGATG
Db      1171 CTGGGGCTGGTCTTGGCCCTATGAAAAGTTCTTGACTTCAGAAAAGCTGTGGGATG
OY      1185 CCATGAAATTTGGTGGCTACGCCACAGCTGAGATCTGAAACAAGAGAGAAACCTGTG
Db      1231 CCATGAAATTTGGTGGCTACGCCACAGCTGAGATCTGAAACAAGAGAGAAACCTGTG
OY      1245 TGGATCTTCTGTCTACCTGAGCTGCTCCAGCAAAATCCATAAGAGGGTTGAGAAGCT
Db      1291 TGGATCTTCTGTCTACCTGAGCTGCTCCAGCAAAATCCATAAGAGGGTTGAGAAGCT
OY      1305 TTGGTCCACCTGGGTGACGCTCAGCTCACTTCTGTGAATGCTTGTGAGCCCTGCTAC
Db      1351 TTGGTCCACCTGGGTGACGCTCAGCTCACTTCTGTGAATGCTTGTGAGCCCTGCTAC
OY      1365 ATCTCTGAGCTTAACTACATGTCTGTGGGTATCACTAGAGTGTGAGTTGTGCACAC
Db      1411 ATCTCTGAGCTTAACTACATGTCTGTGGGTATCACTAGAGTGTGAGTTGTGCACAC
OY      1425 GTGCTCAACGAGAGACTTTGTGTGCTATGCTGTGTGTAGAAAACAGACTGGGAGAC
Db      1471 GTGCTCAACGAGAGACTTTGTGTGCTATGCTGTGTGTAGAAAACAGACTGGGAGAC
OY      1485 TTATGTGAGCAGACATCCACAGAGAAAGGATTTGCTCTTCTTTCTTTCTGTATC
Db      1531 TTATGTGAGCAGACATCCACAGAGAAAGGATTTGCTCTTCTTTCTTTCTGTATC
OY      1545 TTCTGTCTGTGGCAGACTTCAAGACTTTGTGGCTGTGAGGCTATTTAAAGCAGACAG
Db      1591 TTCTGTCTGTGGCAGACTTCAAGACTTTGTGGCTGTGAGGCTATTTAAAGCAGACAG
OY      1605 TTTCAGTGAATTTGATCAATAAAGCTGCTGTCCATCTTGCCCAATGGGAGATGATC
Db      1651 TTTCAGTGAATTTGATCAATAAAGCTGCTGTCCATCTTGCCCAATGGGAGATGATC
OY      1665 TTTCACCAAGAGCTCACCAGCAATTTCCACAGAGATGC
Db      1711 TTTCACCAAGAGCTCACCAGCAATTTCCACAGAGATGC

```

```

PR      11-JAN-2000; 2000JP-0118774.
PR      02-MAY-2000; 2000JP-0183765.
XX
XX
XX      (HELI-) HELIX RES INST.
PI      Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI      Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX      WPI; 2001-524255/58.
XX
XX      830 Primers useful for synthesizing full length cDNA clones and their
PT      use in genetic manipulation.
PS      Claim 2: SEQ ID NO 263; 1380bp + sequence listing; English.
XX
XX      The invention relates to primers for synthesizing full length cDNA
CC      clones. 830 cDNA molecules encoding a human protein have been
CC      isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC      molecules have been determined. Primers for synthesizing the full length
CC      cDNA are useful for clarifying the function of the protein encoded by
CC      the cDNA. The full length clones were obtained by construction of full
CC      length enriched cDNA libraries that were synthesised by the oligo-capping
CC      method. The primers enable the production of the full length cDNA easily
CC      without any special methods. The present sequence is the nucleotide
CC      sequence of the 5'-end of a cDNA provided in the invention.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in CD-ROM format directly from EPO.
XX
SO      Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;
XX
XX      Query Match      34.0%; Score 695; DB 22; Length 877;
XX      Best Local Similarity 100.0%; Pred. No. 0;
XX      Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY      106 CAAGTCTTCCACTTACAGCACAATGCTACTGCTTAAAAAATGAAAGCTGCTGTGTTCT
Db      47 CAAGTCTTCCACTTACAGCACAATGCTACTGCTTAAAAAATGAAAGCTGCTGTGTTCT
OY      166 GGTTCCTCCAGATGGCCATCTTGGCTCTATTTCTTCCACATGTACAGCCACATCAGCTC
Db      107 GGTTCCTCCAGATGGCCATCTTGGCTCTATTTCTTCCACATGTACAGCCACATCAGCTC
OY      226 CCGTCTATGAAGGACACAGCCGAGGAGCAAGCGTGTCTCTCTTCCGGGCTC
Db      167 CCGTCTATGAAGGACACAGCCGAGGAGCAAGCGTGTCTCTCTTCCGGGCTC
OY      286 TGGCTCTTCTTTTGTGGGAGAGCTTTTGGGAGCAGCCAGATGTTTCTTACCTGATGGA
Db      227 TGGCTCTTCTTTTGTGGGAGAGCTTTTGGGAGCAGCCAGATGTTTCTTACCTGATGGA
OY      346 GCCCGCTGGACAGCTGTGATGACCTTCAAGCAGACACCGCTGATGCTGCATGCGC
Db      287 GCCCGCTGGACAGCTGTGATGACCTTCAAGCAGACACCGCTGATGCTGCATGCGC
OY      406 TGTGGGGAGATCGATACGGGCGCTCTTGTGGCGCATGAGAGGCTTTGATGCCATCAT
Db      347 TGTGGGGAGATCGATACGGGCGCTCTTGTGGCGCATGAGAGGCTTTGATGCCATCAT
OY      466 GGAACCTGTGCTCCCGGAGAGACAGTCCAGCTCTTTCACTGGGAGAACAGCCGGGCTGTG
Db      407 GGAACCTGTGCTCCCGGAGAGACAGTCCAGCTCTTTCACTGGGAGAACAGCCGGGCTGTG
OY      526 TTCTGACCTGCTGTCATCATCTCCACACAGATGAATCATCCCGGGGCTCATGCGAG
Db      467 TTCTGACCTGCTGTCATCATCTCCACACAGATGAATCATCCCGGGGCTCATGCGAG
OY      586 GCTCTGTGTCAGTCAACACAGCCCTTTGAGGTGTGGAGAGGCTGCGGCTCTACAGCCA
Db      527 GCTCTGTGTCAGTCAACACAGCCCTTTGAGGTGTGGAGAGGCTGCGGCTCTACAGCCA
OY      646 CGTGTGTCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTGAAGA
Db      587 CGTGTGTCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTGAAGA

```


XX New isolated tumor colon polynucleotide and polypeptide, useful for the
 PT diagnosis, prevention and/or treatment of cancer, in particular colon
 PT cancer

PS Claim 1: SEQ ID NO 2595; 266pp + Sequence Listing; English.

XX The invention relates to a human colon tumour expressed polynucleotide
 CC (I) encoding a polypeptide (II), ABP67991-ABP67996) comprising: (i) any of
 CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
 CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
 CC sequences that hybridize to (i), under moderately stringent conditions;
 CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
 CC degenerate variants of (i). The compositions and methods of the present
 CC invention are useful for the diagnosis, prevention and/or treatment of
 CC cancer, particularly colon cancer. (i) can be used in gene therapy and
 CC (I) and (II) are useful in pharmaceutical compositions such as vaccines.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 505 BP; 120 A; 125 C; 119 G; 141 T; 0 other;

XX Query Match 19.7%; Score 403; DB 24; Length 505;

XX Best Local Similarity 100.0%; Pred. No. 2.6e-189; Mismatches 0; Indels 0; Gaps 0;

XX Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1301 GGCCTTGGCTGACCTGCTGAGCCTTCTGTAATGCTTGTGAGCCTTGC 1360

XX 1 GGCCTTGGCTGACCTGCTGAGCCTTCTGTAATGCTTGTGAGCCTTGC 60

XX 1361 CTACATCTGTGAGCCTTACTACATGCTGTGGGATATCATGAGTGTGATGCTTC 1420

XX 61 CTACATCTGTGAGCCTTACTACATGCTGTGGGATATCATGAGTGTGATGCTTC 120

XX 1421 ACAGTGTGCTCAAGCAAGAGACTTGTGTCATGCTTGTCTGTAATGCTTGTGAGCCTTGC 1480

XX 121 ACAGTGTGCTCAAGCAAGAGACTTGTGTCATGCTTGTCTGTAATGCTTGTGAGCCTTGC 180

XX 1481 AACCTTATGTGAGCAGACATCCACACAGTGAAGAGGATATGCTTCTTCTTCTTCTT 1540

XX 181 AACCTTATGTGAGCAGACATCCACACAGTGAAGAGGATATGCTTCTTCTTCTTCTT 240

XX 1541 GATCTTCTGCTGGGAGACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1600

XX 241 GATCTTCTGCTGGGAGACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 300

XX 1601 ACAGTATGAGTGAATGATTCATTAACCTCCCTGTCACATCTTGGCCCAATGGGAATG 1660

XX 301 ACAGTATGAGTGAATGATTCATTAACCTCCCTGTCACATCTTGGCCCAATGGGAATG 360

XX 1661 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 1703

XX 361 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 403

XX RESULT 8

XX ABR54724

XX ID ABR54724 standard; cDNA; 517 BP.

XX ABR54724;

XX 18-JUN-2002 (first entry)

XX Human colon cancer-associated cDNA, SEQ ID No 194.

XX Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.

XX Homo sapiens.

XX WO200212280-A2.

XX 14-FEB-2002.

XX 30-JUL-2001; 2001WO-US23826.

XX 03-AUG-2000; 2000US-223265P.

XX 02-OCT-2000; 2000US-237406P.

XX 20-MAR-2001; 2001US-277495P.

XX 03-JUL-2001; 2001US-302702P.

XX (CORI-) CORIXA CORP.

XX Pyle RA, Xu J, Secretist H;

XX WPI; 2002-257462/30.

XX Novel polynucleotide encoding colon tumour polypeptides, useful as

XX vaccines for treating colon cancers

XX Claim 1: Page 206; 425pp; English.

XX The invention relates to isolated polynucleotides (I) encoding colon

XX tumour polypeptides (II). (I) is useful for stimulating an immune

XX response in a patient and treating colon cancer in a patient.

XX Oligonucleotides derived from (I) are useful for determining the presence

XX of cancer in a patient. (I) and (II) are useful in pharmaceutical

XX compositions, e.g. vaccines, and other compositions for the diagnosis

XX and treatment of colon cancer. A composition comprising a first component

XX selected from physiologically acceptable carriers and immunostimulants,

XX and an antigen-presenting cell expressing (II) is useful for inhibiting

XX development of cancer in a patient. (I) is useful in the design and

XX preparation of ribozyme molecules for inhibiting expression of tumour

XX polypeptides and (II). ABR5431-ABR55464 represent human colon cancer cDNA

XX sequences of the invention.

XX Sequence 517 BP; 121 A; 130 C; 124 G; 142 T; 0 other;

XX Query Match 19.7%; Score 403; DB 24; Length 517;

XX Best Local Similarity 100.0%; Pred. No. 2.6e-189; Mismatches 0; Indels 0; Gaps 0;

XX Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1301 GGCCTTGGCTGACCTGCTGAGCCTTCTGTAATGCTTGTGAGCCTTGC 1360

XX 13 GGCCTTGGCTGACCTGCTGAGCCTTCTGTAATGCTTGTGAGCCTTGC 72

XX 1361 CTACATCTGTGAGCCTTACTACATGCTGTGGGATATCATGAGTGTGATGCTTC 1420

XX 73 CTACATCTGTGAGCCTTACTACATGCTGTGGGATATCATGAGTGTGATGCTTC 132

XX 1421 ACAGTGTGCTCAAGCAAGAGACTTGTGTCATGCTTGTCTGTAATGCTTGTGAGCCTTGC 1480

XX 133 ACAGTGTGCTCAAGCAAGAGACTTGTGTCATGCTTGTCTGTAATGCTTGTGAGCCTTGC 192

XX 1481 AACCTTATGTGAGCAGACATCCACACAGTGAAGAGGATATGCTTCTTCTTCTTCTT 1540

XX 193 AACCTTATGTGAGCAGACATCCACACAGTGAAGAGGATATGCTTCTTCTTCTTCTT 252

XX 1541 GATCTTCTGCTGGGAGACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1600

XX 253 GATCTTCTGCTGGGAGACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 312

XX 1601 ACAGTATGAGTGAATGATTCATTAACCTCCCTGTCACATCTTGGCCCAATGGGAATG 1660

XX 313 ACAGTATGAGTGAATGATTCATTAACCTCCCTGTCACATCTTGGCCCAATGGGAATG 372

XX 1661 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 1703

XX 373 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 415

XX RESULT 9

XX AAC76156

XX ID AAC76156 standard; cDNA; 2988 BP.

XX AAC76156;

XX 08-FEB-2001 (first entry)
DE Human ORFX ORF1711 polynucleotide sequence SEQ ID NO:3421.
XX
XX Human: open reading frame: ORFX: detection; cytostatic; hepatotropic;
XX valine; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;
XX anticonvulsant; osteoparathic; antiparkinsonian; immunosuppressant; cardiant;
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antineoplastic;
XX antitumor; antibacterial; antifungal; antirheumatic; antihypertensive;
XX antineoplastic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antineoplastic disease; coagulation;
XX thrombosis; contraceptive; ss.
XX
XX Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000MO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Leach M;
XX
XX WPI; 2000-602362/57.
XX P-PSDB; AAB41947.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 5; Page 2597-2599; 5507Pp; English.
XX
XX AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; valine;
XX antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;
XX osteoparathic; anticonvulsant; antiparkinsonian; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antineoplastic; antibacterial; antiviral; antifungal; antirheumatic;
XX antihypertensive; antineoplastic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy. The
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antineoplastic disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 2988 BP; 613 A; 916 C; 842 G; 617 T; 0 other;
XX
XX Query Match 15.0%; Score 306; DB 21; Length 2988;
XX Best Local Similarity 100.0%; Pred. No. 3.7e-141;
XX Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 TTGTATCCATACATGAAACCTGGTCCCGAGAGACAGCCCTCTTCACTGGAGAAC 511
DB 1480 TTGTATCCATACATGAAACCTGGTCCCGAGAGACAGCCCTCTTCACTGGAGAAC 1539
QY 512 AGCCGGGCGCTGTGTTCGACCTGCTGTGACATATCCACAGATGAATCATCC 571
DB 1540 AGCCGGGCGCTGTGTTCGACCTGCTGTGACATATCCACAGATGAATCATCC 1599
QY 572 CGGGCTACATGACAGCTCTGTGACATACAGCCCTTGAAGTGTGGAGAGCCCTGC 631
DB 1600 CGGGCTACATGACAGCTCTGTGACATACAGCCCTTGAAGTGTGGAGAGCCCTGC 1659
QY 632 CGGTCTACAGCCAGCGGTGTCTCAAGAGAGTGGCTTCTCAACCTGACGTCCTAC 691
DB 1660 CGGTCTACAGCCAGCGGTGTCTCAAGAGAGTGGCTTCTCAACCTGACGTCCTAC 1719
QY 692 CGGTCTGAAAGACCCCTCTCAACCTGATATGCTGACCTGTGCGGAGACCCCGG 751
DB 1720 CGGTCTGAAAGACCCCTCTCAACCTGATATGCTGACCTGTGCGGAGACCCCGG 1779
QY 752 GCGCTG 757
DB 1780 GCGCTG 1785

RESULT 10
ID ABR54794 standard; cDNA; 389 BP.
XX ABR54794;
AC ABR54794;
XX 18-JUN-2002 (first entry)
DT
XX Human colon cancer-associated cDNA, SEQ ID NO 264.
DE Human colon cancer; immunogenic; vaccine; tumour; gene; ss.
XX Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.
XX Homo sapiens.
XX
XX WO200212280-A2.
XX
XX 14-FEB-2002.
XX
XX 30-JUL-2001; 2001MO-US23826.
XX
XX 03-AUG-2000; 2000US-223265P.
XX 02-OCT-2000; 2000US-227406P.
XX 20-MAR-2001; 2001US-277495P.
XX 03-JUL-2001; 2001US-302702P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Pyle RA, Xu J, Secrist H;
XX
XX WPI; 2002-257462/30.
XX
XX Novel polynucleotide encoding colon tumour polypeptides, useful as
XX vaccines for treating colon cancers -
XX
XX Claim 1; Page 225; 425Pp; English.
XX
XX The invention relates to isolated polynucleotides (I) encoding colon
XX tumour polypeptides (II). (I) is useful for stimulating an immune
XX response in a patient and treating colon cancer in a patient.
XX Oligonucleotides derived from (I) are useful for determining the presence
XX of cancer in a patient. (I) and (II) are useful in pharmaceutical
XX compositions, e.g. vaccines, and other compositions for the diagnosis
XX and treatment of colon cancer. A composition comprising a first component
XX selected from physiologically acceptable carriers and immunostimulants,
XX and an antigen-presenting cell expressing (II) is useful for inhibiting
XX development of cancer in a patient. (I) is useful in the design and
XX preparation of ribozyme molecules for inhibiting expression of tumour

PS Claim 5; Page 139; 143pp; English.

XX The invention relates to human drug metabolizing enzymes referred as

CC DME and nucleic acid molecules encoding such enzymes. Polynucleotides

CC of the invention are useful for assessing toxicity of test compounds

CC and in gene therapy. Sequences of the invention are useful in the

CC diagnosis, prevention and treatment of autoimmune/inflammatory

CC disorders such as acquired immune deficiency syndrome (AIDS), adult

CC respiratory distress syndrome, allergies, anaemia, atherosclerosis,

CC asthma, autoimmune haemolytic anaemia, contact dermatitis, Crohn's

CC disease, glomerulonephritis, Goodpasture's syndrome, gout, Grave's

CC disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple

CC sclerosis, osteoarthritis, osteoporosis, psoriasis, systemic lupus

CC erythematosus, rheumatoid arthritis, scleroderma, ulcerative colitis,

CC uveitis, viral, bacterial, fungal, parasitic, protozoal, helminthic

CC infections; cell proliferative disorders such as actinic keratosis,

CC arteriosclerosis, atherosclerosis, Duchenne and Becker dystrophy,

CC hepatitis, Cushing's syndrome, cancers, myelodysplastic syndrome,

CC epilepsy; endocrine disorders such as disorders of the hypothalamus

CC and pituitary resulting from lesions such as primary brain tumours,

CC adenomas, infarction associated with pregnancy, aneurysms, vascular

CC malformations; eye disorders such as conjunctivitis, iritis, retinitis,

CC glaucoma, pigmentosa, metabolic disorders such as Addison's disease,

CC cystic fibrosis, diabetes, goitre, glycogen storage diseases,

CC hypercholesterolaemia, hyperthyroidism, hypoglycaemia, lipid myopathies,

CC Menkes syndrome, mannosidosis, obesity; gastrointestinal disorders such

CC as dysphagia, gastric carcinoma, anorexia, nausea, gastroenteritis,

CC hyperillirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's

CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental

CC disorders. The present sequence is human DME-5 protein cDNA.

XX

SQ Sequence 1647 BP; 287 A; 560 C; 501 G; 299 T; 0 other;

Query Match 2.9%; Score 59; DB 24; Length 1647;

Best Local Similarity 100.0%; Pred. No. 1.4e-18;

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 624 AGGCGTCCGCTCTACAGCCAGCGTGCTCAAGAGAGTGCGCTTCTCAACCTGCAG 682

DB 488 AGGCGTCCGCTCTACAGCCAGCGTGCTCAAGAGAGTGCGCTTCTCAACCTGCAG 546

RESULT 13

AAD02700

ID AAD02700 standard; cDNA; 1694 BP.

AC AAD02700;

XX

DT 02-MAY-2001 (first entry)

XX

DE Human glycosyl sulfotransferase-4beta (GST-4beta) cDNA.

XX

Human: glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive;

XX therapy; selectin binding inhibitor; gene therapy; inflammation;

XX systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;

XX polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;

XX glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailitis;

XX Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;

XX demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;

XX myocarditis; adult respiratory distress syndrome; eczema; psoriasis;

XX asthma; hypersensitivity; rheumatic fever; tissue rejection;

XX chromosome 16q23.1; ss.

XX

OS Homo sapiens.

XX

XX

Key Location/Qualifiers

FT 5'UTR 9..188

FT CDS 189..1376

FT /tag- a

FT /tag- b

FT /product- "Human glycosyl sulfotransferase-4beta

FT (GST-4beta)"

FT /note- "CDS is referred as SEQ ID NO:21 in brief

FT description of the figures (page no.: 4)"

FT 3'UTR 1377..1694

FT /tag- c

XX

XX WO200106015-A1.

XX

XX 25-JAN-2001.

XX

XX 19-JUL-2000; 2000WO-US19741.

XX

XX 20-JUL-1999; 99US-0144694.

XX

XX 13-JUL-2000; 2000US-0593828.

XX

XX (REGC) UNIV CALIFORNIA.

XX

XX Rosen SD, Lee JK, Hemmerich S;

XX

XX WPI; 2001-138471/14.

XX

XX P-PSDB; NAY72640.

XX

XX

XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for

XX diagnostic and therapeutic agent screening applications

XX

XX Claim 6; Fig 4A; 128pp; English.

XX

XX The present sequence is human glycosyl sulfotransferase-4beta (GST-4

XX beta) cDNA. GST-4 gene is found on chromosome 16q23.1.

XX GST is a type 2 membrane protein useful for inhibiting a binding event

XX between a selectin and a selectin ligand, which comprises contacting the

XX agent that inhibits the sulphation activity of GST. GST is also useful

XX in inhibiting a selectin mediated binding event. GST is useful in gene

XX therapy to treat disorders such as acute or chronic inflammation,

XX systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis

XX nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,

XX glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's

XX disease, Grave's disease, adrenailitis, hypoparathyroidism, pernicious

XX anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,

XX dermatitis, myocarditis, regional enteritis, adult respiratory distress

XX syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,

XX bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection

XX during transplantation.

XX

SQ Sequence 1694 BP; 259 A; 586 C; 514 G; 335 T; 0 other;

Query Match 2.9%; Score 59; DB 22; Length 1694;

Best Local Similarity 100.0%; Pred. No. 1.4e-18;

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 624 AGGCGTCCGCTCTACAGCCAGCGTGCTCAAGAGAGTGCGCTTCTCAACCTGCAG 682

DB 676 AGGCGTCCGCTCTACAGCCAGCGTGCTCAAGAGAGTGCGCTTCTCAACCTGCAG 734

RESULT 14

AAD02699

ID AAD02699 standard; cDNA; 2044 BP.

AC AAD02699;

XX

DT 02-MAY-2001 (first entry)

XX

DE Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #2.

XX

Human: glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;

XX therapy; selectin binding inhibitor; gene therapy; inflammation;

XX systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;

XX polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;

XX glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailitis;

XX Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;

XX demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;

XX myocarditis; adult respiratory distress syndrome; eczema; psoriasis;

XX asthma; hypersensitivity; rheumatic fever; tissue rejection;

.KW chromosome 16q23.1; ss.
.XX
OS Homo sapiens.
.XX Key Location/Qualifiers
FH CDS 218..1390
FT /tag-a
FT /product- "Human glycosyl sulfotransferase-4alpha
FT /GST-4alpha"
FT /note- "CDS is specifically claimed as SEQ ID NO: 4
in claim 6 (page no: 41) of the specification."
FT
TN MO200106015-A1.
PN
PD 25-JAN-2001.
RX
TX 19-JUL-2000; 2000MO-US19741:
PR 20-JUL-1999; 99US-0144694.
PR 13-JUL-2000; 2000US-0593828.
PA (REGC) UNIV CALIFORNIA.
PI Rosen SD, Lee JK, Hemmerich S;
DR WPI: 2001-138471/14.
DX P-PADB; AAY72639.
PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT diagnostic and therapeutic agent screening applications -
XX Claim 6; Fig 1; 128pp; English.

The present sequence is human glycosyl sulfotransferase-4alpha (GST-4 alpha) cDNA. GST-4 gene is found on chromosome 16q23.1.
GST is a type 2 membrane protein useful for inhibiting a binding event between a selectin and a selectin ligand, which comprises contacting the selectin with a non-sulphated selectin ligand. GST and a small molecular agent that inhibits the sulphation activity of GST. GST is also useful in inhibiting a selectin mediated binding event. GST is useful in gene therapy to treat disorders such as acute or chronic inflammation,
systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes, glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's disease, Graves' disease, adenomatous hyperplasia, hypoparathyroidism, pernicious anemia, demyelinating diseases, cirrhosis, ulcerative colitis, dermatitis, myocarditis, regional enteritis, adult respiratory distress syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis, bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection during transplantation.
Note: The present sequence is also shown in sequence listing (page no: 56) but lacks four nucleotides at its 3' end.

Sequence 2044 BP; 422 A; 645 C; 570 G; 407 T; 0 other;

Query Match 2.9%; Score 59; DB 22; Length 2044;
Best Local Similarity 100.0%; Pred. No. 1.4e+18;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps

Gy 624 AGGCCTCCCGCTCCTACAGCCAGGTGGTGCTCAAGAAGTGGCTTTCAAACCTGCAG 682
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 708 AGGCCCTCCCGCTCCTACAGCCAGGTGGTGCTCAAGAAGTGGCTTTCAAACCTGCAG 766

RESULT 15
ID ADD02698 standard; cDNA: 2170 BP.
XX ADD02698;
XX AC
DT 02-MAY-2001 (first entry)
DE Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #1.

XX	Human: glycosyl sulfotransferase-4alpha: GST-4alpha; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation; systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes; polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis; glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis; Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia; demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis; myocarditis; adult respiratory distress syndrome; eczema; psoriasis; asthma; hypersensitivity; Rheumatic fever; tissue rejection; chromosome 10q23.1; ss.
XX	Homo sapiens.
XX	Key
XX	Location/Qualifiers
XX	5'UTR
XX	9..343
XX	/tag= a
XX	344..1516
XX	CDS
XX	/tag= b
XX	/product= "Human glycosyl sulfotransferase-4alpha
XX	(GST-4alpha)"
XX	/note= "CDS is specifically claimed as SEQ ID NO: 4
XX	in claim 6 (page no: 41) of the specification"
XX	1517..2134
XX	3'UTR
XX	/tag= c
XX	WO200106015-A1.
XX	25-JAN-2001.
XX	19-JUL-2000; 2000WO-US19741.
XX	20-JUL-1999; 99US-0144694.
XX	13-JUL-2000; 2000US-0593828.
XX	(REGC) UNIV CALIFORNIA.
XX	Rosen SD, Lee JK, Hemmerich S;
XX	WPI: 2001-138471/14.
XX	P-PSDB; MAY72639.
XX	New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for diagnostic and therapeutic agent screening applications
XX	Claim 6; Page 62; 128pp; English.
XX	The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
XX	alpha) cDNA. GST-4 gene is found on chromosome 16q23.1.
XX	GST is a type 2 membrane protein useful for inhibiting a binding event
XX	between a selectin and a selectin ligand, which comprises contacting the
XX	selectin with a non-sulphated selectin ligand, GST and a small molecular
XX	agent that inhibits the sulphation activity of GST. GST is also useful
XX	in inhibiting a selectin mediated binding event. GST is useful in gene
XX	therapy to treat disorders such as acute or chronic inflammation,
XX	systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
XX	nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
XX	glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
XX	disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
XX	anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
XX	dermatitis, myocarditis, regional enteritis, adult respiratory distress
XX	syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
XX	bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
XX	during transplantation.
XX	Sequence 2170 BP; 449 A; 682 C; 612 G; 427 T; 0 other:
XX	Query Match 2.9%; Score 59; DB 22; Length 2170;
XX	Best Local Similarity 100.0%; Pred. No. 1.4e-18;
XX	Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

• Fri Aug 15 08:09:33 2003

us-09-645-078-1_3.rng

Page 13

Db 834 AGGCTGCGGCTCCTACAGCCACGTGGTGTCAAGAGAGTGGCTTCTTCACCTGCAG 892

Search completed: August 14, 2003, 18:09:57
Job time : 544 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 15:34:51 ; Search time 7280 Seconds

(without alignments)
11480.545 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043

Sequence: 1 gaattcattgtgtggtta.....tgggatacctataaaaaaaa 2043

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5770842

Minimum DB seq length: 10

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:

1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_din:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcg_hum:*
40: em_hcg_mus:*
41: em_hcg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1722	84.3	1992	9	AF280088 Homo sapi
2	1653	80.9	2037	9	BC035282 Homo sapi
3	1598	78.2	2011	9	AK026635 Homo sapi
4	1597	78.2	183228	2	AC010547 Homo sapi
5	1597	78.2	206943	2	AC138848 Homo sapi
6	1547	75.7	1979	6	BD127258 Homo sapi
7	1547	75.7	1979	6	AK074746 Homo sapi
8	1547	75.7	2032	6	AR203335 Sequence
9	1547	75.7	2032	6	BD134772 Glycosyls
10	1547	75.7	2032	6	AF131235 Homo sapi
11	1547	75.7	2032	6	AF149783 Homo sapi
12	1547	75.7	2032	6	AR310485 Sequence
13	1547	75.7	2032	6	BD124832 Primer fo
14	1547	75.7	2032	6	BD126950 primer fo
15	1547	75.7	2032	6	AX381256 Sequence
16	1547	75.7	2032	6	AX381326 Sequence
17	1547	75.7	2032	6	BD125617 Primer fo
18	1547	75.7	2032	6	AF176839 Homo sapi
19	1547	75.7	2032	6	AX327330 Sequence
20	1547	75.7	2032	6	AF176838 Homo sapi
21	1547	75.7	2032	6	AF219990 Homo sapi
22	1547	75.7	2032	6	AF246718 Homo sapi
23	1547	75.7	2032	6	AF280086 Homo sapi
24	1547	75.7	2032	6	AF219991 Homo sapi
25	1547	75.7	2032	6	AC009163 Homo sapi
26	1547	75.7	2032	6	AC025287 Homo sapi
27	1547	75.7	2032	6	AC009105 Homo sapi
28	1547	75.7	2032	6	AR203340 Sequence
29	1547	75.7	2032	6	BD134774 Glycosyls
30	1547	75.7	2032	6	BD134553 Method fo
31	1547	75.7	2032	6	BD134773 Glycosyls
32	1547	75.7	2032	6	AF131236 Mus muscu
33	1547	75.7	2032	6	AF109155 Mus muscu
34	1547	75.7	2032	6	AC130174 Rattus no
35	1547	75.7	2032	6	AC110314 Rattus no
36	1547	75.7	2032	6	AC121740 Rattus no
37	1547	75.7	2032	6	AC133449 Rattus no
38	1547	75.7	2032	6	AC096328 Rattus no
39	1547	75.7	2032	6	AC095651 Rattus no
40	1547	75.7	2032	6	AC137365 Rattus no
41	1547	75.7	2032	6	AC129764 Rattus no
42	1547	75.7	2032	6	AC094919 Rattus no
43	1547	75.7	2032	6	AF176841 Mus muscu
44	1547	75.7	2032	6	AF176840 Mus muscu
45	1547	75.7	2032	6	AC099262_1 Continuation (2 of

ALIGNMENTS

RESULT 1
AF280088
LOCUS Homo sapiens l-selectin ligand sulfotransferase GST-3 mRNA, complete cds.
ACCESSION AF280088
VERSION AF280088.1 GI:12060807
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1992)
Hemmerich,S., Lee,J.K., Bhakta,S., Bistrup,A., Ruddle,N.R. and Rosen,S.D.

TITLE Chromosomal localization and genomic organization for the
galactose/N-acetylglucosamine/N-acetylglucosamine
6-O-sulfotransferase gene family
JOURNAL Glycobiology 11 (1), 75-87 (2001)
MEDLINE 21096027
PUBMED 11181564
REFERENCE 2 (bases 1 to 1992)
AUTHORS Hemmerich, S., Bhakta, S., Lee, J.-K., Bistrup, A., Ruddie, N.R. and
Rosen, S.D.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience,
3401 Hillview Ave., Palo Alto, CA 94304, USA
FEATURES
source
1. 1992
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q23.1-q23.2"
/cell_type="high endothelial"
/tissue_type="tonsil"
1. 1992
/gene="GST3"
132. 1282
/gene="GST3"
/note="similar to Homo sapiens sequence encoded by GenBank
Accession Number AF131235"
/codon_start=1
/product="L-selectin ligand sulfotransferase GST-3"
/protein_id="AF131235.1"
/db_xref="GI:12060808"
/translation="MLPKKMLLEFLVQMALILFFHMYSHNISLSKQKOPRHH
VLVLSMRGSSSFVGLFGQHPDVFELMEPMHVMTEKQSTAMLMHMAVRDILRAVE
LQDMSEVDAIMERPGRSSFLQPMENRALCSNPACDIIPDEIIPRACRLSCQOP
FEVFEKACRSYSHVLEKVEFNLQSLYDLKDPISLHLIIVHLPRAVFRSRETK
GELMIDRSIVMGOHEOKLEKEDOPYVVMVICOSELEIKTISLPRALOEKRLVRY
EDLAPVPAQTSRMEFVGLLEFLPHLOTQVHNTTRGKGMDHAFHNADALNVQAM
RMSLPREKYSRQKAGDAMNLTGRHVHSEDOQRLLDLSTVTPEDIH"

BASE COUNT 482 a 540 c 489 g 481 t
ORIGIN

Query Match 84.3%; Score 1722; DB 9; Length 1992;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1722; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CATTGTTGGTACGGAAGACGACAGAAAGGTAGAGAGAAAAGCGCATGCCGCGTA 66
b 1 CATTGTTGGTACGGAAGACGACAGAAAGGTAGAGAGAAAAGCGCATGCCGCGTA 60
QY 67 GCAGTAGGCTCTCAAAAGCAGAGGAGGAGCCCAAGCCCAAGTCTTCCACTTCAGCAC 126
Db 61 GCAGTAGGCTCTCAAAAGCAGAGGAGGAGCCCAAGCCCAAGTCTTCCACTTCAGCAC 120
QY 127 AATGCTACTGCTCAAAAATGAAGCTCGTGTGTTGTTGCCAGATGCGCATCT 186
Db 122 AATGCTACTGCTCAAAAATGAAGCTCGTGTGTTGTTGCCAGATGCGCATCT 180
QY 187 GCGCTATTCTTCACATGTACAGCCCAACATCAGCTCCCTGTATGAAGGACAGCC 246
Db 181 GCGCTATTCTTCACATGTACAGCCCAACATCAGCTCCCTGTATGAAGGACAGCC 240
QY 247 CGAGGCATGCAAGCTGCTGTTCTGTTCTGCTGCGCTCTGCTCTTTTGGGGCA 306
Db 241 CGAGGCATGCAAGCTGCTGTTCTGTTCTGCTGCGCTCTGCTCTTTTGGGGCA 300
QY 307 GCTTTTGGGACAGCCAGATGTTTCTACCTGATGAGAGCCGCTGGAGCTGGAT 366
Db 301 GCTTTTGGGACAGCCAGATGTTTCTACCTGATGAGAGCCGCTGGAGCTGGAT 360
QY 367 GACCTTCAAGCAGAGCAGCGCTGATGCTGACATGCTGCGGATCTGATACGGC 426
Db 361 GACCTTCAAGCAGAGCAGCGCTGATGCTGACATGCTGCGGATCTGATACGGC 420

QY 427 CGTCTCTTGTGCGATGAGCGCTCTTGTGATGCTCATGGAACCTGTCCCGAGACA 486
Db 421 CGTCTCTTGTGCGATGAGCGCTCTTGTGATGCTCATGGAACCTGTCCCGAGACA 480
QY 487 GTCCAGCCCTTTCATGAGGAGAAACAGCCGGGCGCTGTCTGACACCTGCTGACAT 546
Db 481 GTCCAGCCCTTTCATGAGGAGAAACAGCCGGGCGCTGTCTGACACCTGCTGACAT 540
QY 547 CATCCCAAGATGATATATATCCCGGGGCTCACTCAGAGCTCTGACATCAAGCC 606
Db 541 CATCCCAAGATGATATATATCCCGGGGCTCACTCAGAGCTCTGACATCAAGCC 600
QY 607 CTTTGAAGTGTGAGAAAGGCTGCGCTCTACAGCCACAGTGTCTCAAGAGGTGCG 666
Db 601 CTTTGAAGTGTGAGAAAGGCTGCGCTCTACAGCCACAGTGTCTCAAGAGGTGCG 660
QY 667 CTTCTTCAACCTGACATCCCTCTACCGGCTGCGGAGAAAGCCCTCCCTCACTGATAT 726
Db 661 CTTCTTCAACCTGACATCCCTCTACCGGCTGCGGAGAAAGCCCTCCCTCACTGATAT 720
QY 727 CGTGAACCTGTCGGGAGACCCCGGGCGGTTCGCTCCGAGAAAGCAGAAAGGAGA 786
Db 721 CGTGAACCTGTCGGGAGACCCCGGGCGGTTCGCTCCGAGAAAGCAGAAAGGAGA 780
QY 787 TCTCATGATTGACAGTCCGATTGTATGGGCGACATGACAAAACCTCAAGAGAGA 846
Db 781 TCTCATGATTGACAGTCCGATTGTATGGGCGACATGACAAAACCTCAAGAGAGA 840
QY 847 CCACACCTATGATGATGAGAGTCAATGCGCCAAACCCAGCTGAGATCTCAAGACAT 906
Db 841 CCACACCTATGATGATGAGAGTCAATGCGCCAAACCCAGCTGAGATCTCAAGACAT 900
QY 907 CCAGTCTTGGCCCAAGGCGCTGAGAGACGTAACGCTTGTGCGTATGAGACCTGCG 966
Db 901 CCAGTCTTGGCCCAAGGCGCTGAGAGACGTAACGCTTGTGCGTATGAGACCTGCG 960
QY 967 TCGAGCCCTGTGCGCCAGACCTCCGAATGTATGAAATCTGCGGAAATTCCTGCC 1026
Db 961 TCGAGCCCTGTGCGCCAGACCTCCGAATGTATGAAATCTGCGGAAATTCCTGCC 1020
QY 1027 CCATCTTACAGACCTGGGTGATATCATCAACCCGAGCAAGGCGATGGTACACGCTTT 1086
Db 1021 CCATCTTACAGACCTGGGTGATATCATCAACCCGAGCAAGGCGATGGTACACGCTTT 1080
QY 1087 CCACACAATGCCAGGATGCTTATATGTCTCCAGGCTTGCGCTGCTTGGCCTTA 1146
Db 1081 CCACACAATGCCAGGATGCTTATATGTCTCCAGGCTTGCGCTGCTTGGCCTTA 1140
QY 1147 TGAAGAGTTTTCGACTTCAGAAAGCTGTGCGATGCGAATTTGCTGGGCTAACCG 1206
Db 1141 TGAAGAGTTTTCGACTTCAGAAAGCTGTGCGATGCGAATTTGCTGGGCTAACCG 1200
QY 1207 CCAGCTCAGATCGAACAAGAACAGAGAAAACCTGTGTGATCTTCTCTACCTGAC 1266
Db 1201 CCAGCTCAGATCGAACAAGAACAGAGAAAACCTGTGTGATCTTCTCTACCTGAC 1260
QY 1267 TGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAAAGCTTGTGCCACCTGTGTACGC 1326
Db 1261 TGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAAAGCTTGTGCCACCTGTGTACGC 1320
QY 1327 TCACTACCTTTCCTGAATGCTCTGAGGCTTCCCAACATCTGTGAGGCTTAACCTACATG 1386
Db 1321 TCACTACCTTTCCTGAATGCTCTGAGGCTTCCCAACATCTGTGAGGCTTAACCTACATG 1380
QY 1387 TCTGTGGGTATCACATGATGATGATGTTGTGTCACACGTGCTCAAGCAGAAAGACTTTT 1446
Db 1381 TCTGTGGGTATCACATGATGATGATGTTGTGTCACACGTGCTCAAGCAGAAAGACTTTT 1440
QY 1447 GTGTCCATGCTGTGTCTAGAAAAACAGACTGGGGAACCTTATGTGAGACACATCCAC 1506
Db 1441 GTGTCCATGCTGTGTCTAGAAAAACAGACTGGGGAACCTTATGTGAGACACATCCAC 1500
QY 1507 CAGTGAACAAGGATATGCTCTTCTTCTTCTGATCTTCTGCTGGGACAGACTTTCAG 1566

BASE COUNT	507 a	546 c	481 g	503 t	
ORIGIN					
Query Match	80.9%;	Score 1653;	DB 9;	Length 2037;	
Best Local Similarity	100.0%;	Prod. No. 0;			
Matches 1653;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
OY	51	GCSCATGGCCCGCGGTACGATGAGCCCTCTCAAAAGCAGCAGGAAAGCCCAAGCACAAGG	110		
Db	1	GCSCATGGCCCGCGGTACGATGAGCCCTCTCAAAAGCAGCAGGAAAGCCCAAGCACAAGG	60		
OY	111	TCCTCCACTTCAGACAAATGCTACTGGCTAAAAAATGAAGCTCCTGCTGTTTGGGTTT	170		
Db	61	TCCTCCACTTCAGACAAATGCTACTGGCTAAAAAATGAAGCTCCTGCTGTTTGGGTTT	120		
OY	171	CCGAGATGGCCATCTTGGCTCTATTCTTCCACATGTACAGCCACACAATCAGTCCCTGT	230		
Db	121	CCGAGATGGCCATCTTGGCTCTATTCTTCCACATGTACAGCCACACAATCAGTCCCTGT	180		
OY	231	CTATGAAGGCACAGCCCGAGGCGATGCAAGTGTGTGTCTGTCTCTGGCGCTGTGGCT	290		
Db	181	CTATGAAGGCACAGCCCGAGGCGATGCAAGTGTGTGTCTGTCTCTGGCGCTGTGGCT	240		
OY	291	CTTCTTTTGTGTGGGCGAGCTTTTGTGGGACAGCCACAGATGTTTCTACTGATGGAAGCCG	350		
Db	241	CTTCTTTTGTGTGGGCGAGCTTTTGTGGGACAGCCACAGATGTTTCTACTGATGGAAGCCG	300		
OY	351	CCTGGCACAGTGTGATGACTTTCAGACGACAGCACCGCGCTGGATGCTCACATGGCTGTGC	410		
Db	301	CCTGGCACAGTGTGATGACTTTCAGACGACAGCACCGCGCTGGATGCTCACATGGCTGTGC	360		
OY	411	GGGATCTGATACGGGCGCTCTCTTGTGTGCACATGAGCGTCTTGTGATGCTTACATGGAAC	470		
Db	361	GGGATCTGATACGGGCGCTCTCTTGTGTGCACATGAGCGTCTTGTGATGCTTACATGGAAC	420		
OY	471	CTGGTCCCCGGAGACAGTCCAGCCTCTTTCAGTGTGGGAAACAGCCGGGCGCTGTGTTCTG	530		
Db	421	CTGGTCCCCGGAGAGACAGTCCAGCCTCTTTCAGTGTGGGAAACAGCCGGGCGCTGTGTTCTG	480		
OY	531	CACCTGCCCTGTGAATCATCCACACAAGATGAATCATCCCGCGGCTCACTGCAGGCTCC	590		
Db	481	CACCTGCCCTGTGAATCATCCACACAAGATGAATCATCCCGCGGCTCACTGCAGGCTCC	540		
OY	591	TGTGCAGTCAACAGCCCTTTGAGGTGTGTGAAAGGCGCTGCCGCTCTACAGCCAGCTGG	650		
Db	541	TGTGCAGTCAACAGCCCTTTGAGGTGTGTGAAAGGCGCTGCCGCTCTACAGCCAGCTGG	600		
OY	651	TGCTCAAGAGAGGTGCGCTTCTTCAACCTGCAGTCCCTCTACCCGCTGCTGAAAGACCCCT	710		
Db	601	TGCTCAAGAGAGTGCCTTCTTCAACCTGCAGTCCCTCTACCCGCTGCTGAAAGACCCCT	660		
OY	711	CCGTCACCTCATATGCTGTGCAACCTGGTCCGGAGACCCCGGCGCGCTGTGCTGCCGAG	770		
Db	661	CCGTCACCTCATATGCTGTGCAACCTGGTCCGGAGACCCCGGCGCGCTGTGCTGCCGAG	720		
OY	771	AACGCACAAAGGAGATCTCATGATTGACAGTGCATTTGTGATGGGGCAGCATGAGCAAA	830		
Db	721	AACGCACAAAGGAGATCTCATGATTGACAGTGCATTTGTGATGGGGCAGCATGAGCAAA	780		
OY	831	AACCTCAAGAGAGAGACCAACCTCACTATGTGTGATGCAAGTCAATGTGCCAAAGCAGCTGG	890		
Db	781	AACCTCAAGAGAGAGACCAACCTCACTATGTGTGATGCAAGTCAATGTGCCAAAGCAGCTGG	840		
OY	891	AGATCTCAACAAGCATCAGTCTTGGCCAAAGGCGCTGCAGGAACGCTACTGCTTGTGC	950		
Db	841	AGATCTCAACAAGCATCAGTCTTGGCCAAAGGCGCTGCAGGAACGCTACTGCTTGTGC	900		

```

QY 951 GCTATAGAGACCTGGCTCGAGCCCTGTGGCCAGACTTCCGATGATGAATTCGTGG 1010
Db 901 GCTATAGAGACCTGGCTCGAGCCCTGTGGCCAGACTTCCGATGATGAATTCGTGG 960
QY 1011 GATTGGAATTTCTGCCCCATCTTCAGACCTGGGTGCATTAACATCACCAGGGAAGGCA 1070
Db 961 GATTGGAATTTCTGCCCCATCTTCAGACCTGGGTGCATTAACATCACCAGGGAAGGCA 1020
QY 1071 TGGGTGACCAAGCTTTTCACACAAATGCCAGGAGTCCCTTAATGTCTCCAGGCTTGGC 1130
Db 1021 TGGGTGACCAAGCTTTTCACACAAATGCCAGGAGTCCCTTAATGTCTCCAGGCTTGGC 1080
QY 1131 GCGGCTTGTGCCCCATCTTCAGAAAGCTTTCAGATTCAGAAAGCTTGGGAGTCCATGA 1190
Db 1081 GCGGCTTGTGCCCCATCTTCAGAAAGCTTTCAGATTCAGAAAGCTTGGGAGTCCATGA 1140
QY 1191 ATTGCTGGGCTACCCGACAGCTGATGATGTAACAAAGAGAGAAACCTTGTCTGGATC 1250
Db 1141 ATTGCTGGGCTACCCGACAGCTGATGATGTAACAAAGAGAGAAACCTTGTCTGGATC 1200
QY 1251 TTCTGCTACCTGGAGCTGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAAGGCTTGTG 1310
Db 1201 TTCTGCTACCTGGAGCTGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAAGGCTTGTG 1260
QY 1311 CCACCTGGTGGAGCTGACATCTTCTGTGAATGCTTGTGAGCTTGGCTTACATCTCT 1370
Db 1261 CCACCTGGTGGAGCTGACATCTTCTGTGAATGCTTGTGAGCTTGGCTTACATCTCT 1320
QY 1371 GAGCCTTAACATCATCTGTGTGGGTATCAACACTAGATGTGAGTGTGTCACACAGCTTC 1430
Db 1321 GAGCCTTAACATCATCTGTGTGGGTATCAACACTAGATGTGAGTGTGTCACACAGCTTC 1380
QY 1431 AAGCACAAGAGACTTTGTGTGCTCATGCTTGTGTAGAAACAGACTGGGGAACCTTATGT 1490
Db 1381 AAGCACAAGAGACTTTGTGTGCTCATGCTTGTGTAGAAACAGACTGGGGAACCTTATGT 1440
QY 1491 GAGCAGCAGATCCACAGCAAGGAAAGGATGCTTCTTCTTGTGATCTTCCG 1550
Db 1441 GAGCAGCAGATCCACAGCAAGGAAAGGATGCTTCTTCTTGTGATCTTCCG 1500
QY 1551 TCTGGGACAGACTTCAGAGACTTGTGGCTGAGAGGCTTATTAAGCAGACAGATATCAG 1610
Db 1501 TCTGGGACAGACTTCAGAGACTTGTGGCTGAGAGGCTTATTAAGCAGACAGATATCAG 1560
QY 1611 TGGATTTGATCCATTAACCTCCCTGTCCACATCTTCCCAATGAGGAAATGATCTTTCAC 1670
Db 1561 TGGATTTGATCCATTAACCTCCCTGTCCACATCTTCCCAATGAGGAAATGATCTTTCAC 1620
QY 1671 CAAGAGCTCAGCAGATTTTCCACAGAGATGC 1703
Db 1621 CAAGAGCTCAGCAGATTTTCCACAGAGATGC 1653

RESULT 3
AK026635
LOCUS Homo sapiens cDNA, FLJ22982 f1s, clone KAT11454, highly similar to
DEFINITION AF131235 Homo sapiens N-acetylglucosamine 6-O-sulfotransferase
mRNA.
ACCESSION AK026635
VERSION AK026635.1 GI:10439531
KEYWORDS oligo capping; f1s (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (sities)
Matsumoto, K., Kumagai, A., Itakura, S., Yamazaki, M., Tashiro, H.,
Ota, T., Suzuki, Y., Ohashi, M., Nishitani, T., Shibahara, T., Tanaka, T.,
Nakamura, Y., Isogai, T. and Sugano, S.
NEBO human cDNA sequencing project
Unpublished
REFERENCE 2 (bases 1 to 2011)

```

```

AUTHORS Sugano, S., Suzuki, Y., Ota, T., Ohashi, M., Nishitani, T., Isogai, T.,
TITLE Shibahara, T., Tanaka, T. and Nakamura, Y.
JOURNAL Direct Submission
Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
Japan (E-mail: cdna@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286,
Fax: 81-3-5449-5416)
COMMENT NEBO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing: Research Association for Biotechnology: cDNA library
construction, 5'-3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
Location/Qualifiers
1..2011
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="KAT11454"
/cell_line="KATO III"
/cell_type="signet-ring cell carcinoma"
/clone_id="KAT"
/notes="cloning vector pME18SFL3"
1..2011
/notes="highly similar to AF131235 Homo sapiens
N-acetylglucosamine 6-O-sulfotransferase mRNA"
BASE COUNT 482 a 541 c 475 g 513 t
ORIGIN
Query Match 78.2%; Score 1598; DB 9; Length 2011;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1598; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 106 CAAGGCTTCCACTTACAGCAATGCTACTGCTTAAATGAAGAACTCTGCTGTTCT 165
Db 50 CAAGGCTTCCACTTACAGCAATGCTACTGCTTAAATGAAGAACTCTGCTGTTCT 109
QY 166 GGTTCACAGAGGAGCATTTGGCTATCTTCCACATGATGAGCACAATCAGCTC 225
Db 110 GGTTCACAGAGGAGCATTTGGCTATCTTCCACATGATGAGCACAATCAGCTC 169
QY 226 CCTGTCTATGAAGGACAGCCGAGGCGATGCAAGTGTGTCTCTCTGCGGCTC 285
Db 170 CCTGTCTATGAAGGACAGCCGAGGCGATGCAAGTGTGTCTCTCTGCGGCTC 229
QY 286 TGGCTCTCTTTGTGGGAGAGCTTTTGGGAGCAGCAGATGTTTCTACTGATGGA 345
Db 230 TGGCTCTCTTTGTGGGAGAGCTTTTGGGAGCAGCAGATGTTTCTACTGATGGA 289
QY 346 GCCCGCTGGCAGCTGTGATGACCTTCAAGCAGAGACAGCGCTGATCTGCATGCG 405
Db 290 GCCCGCTGGCAGCTGTGATGACCTTCAAGCAGAGACAGCGCTGATCTGCATGCG 349
QY 406 TGTGGGAGATCTGATACGCGGCTCTTCTTGTGCGACATGAGCGCTTGTGATCT 465
Db 350 TGTGGGAGATCTGATACGCGGCTCTTCTTGTGCGACATGAGCGCTTGTGATCT 409
QY 466 GGAACCTGTGCCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAACAGCGCGGCTG 525
Db 410 GGAACCTGTGCCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAACAGCGCGGCTG 469
QY 526 TTCTGACCTGCTGTGATCATCTCCACAAATGATCAATCCCGGCTCAGTCTGAG 585
Db 470 TTCTGACCTGCTGTGATCATCTCCACAAATGATCAATCCCGGCTCAGTCTGAG 529
QY 586 GCTCTGTGACATCAACAGCCCTTGTGAGGTGTGAGAAAGGCTCGCTCTTACAGCA 645
Db 530 GCTCTGTGACATCAACAGCCCTTGTGAGGTGTGAGAAAGGCTCGCTCTTACAGCA 589
QY 646 CGTGTGCTCAAGAGGAGTGGCTTCTTCAACTGAGTCCCTTACCGGCTCTGAAAGA 705

```


Db 590 CTTGTCCTCAGAGAGGTGGCTTCTTCAACCTGACAGTCCCTCTACCCGCTGTGAAAGA 649
 QY 706 CCCCTGCTCAACCTGATATCGNCACTGCTGCGGAGACCCCGGCGCTTCCCTTC 765
 Db 650 CCCCTGCTCAACCTGATATCGNCACTGCTGCGGAGACCCCGGCGCTTCCCTTC 709
 QY 766 CCGAAGACGACAAAGGAGATCTCATGATTTGACAGTCCGATTTGATGGGAGCATGA 825
 Db 710 CCGAAGACGACAAAGGAGATCTCATGATTTGACAGTCCGATTTGATGGGAGCATGA 769
 QY 826 GCAAAAACCTCAAGAGAGACACCACTTATGATGATGAGTCAATCTGCAAGCA 885
 Db 770 GCAAAAACCTCAAGAGAGACACCACTTATGATGATGAGTCAATCTGCAAGCA 829
 QY 886 GCTGAGATCTACAAGACCATCCAGTCTTGGCCCAAGGCTTGAGAGAGCTTCCCTCT 945
 Db 830 GCTGAGATCTACAAGACCATCCAGTCTTGGCCCAAGGCTTGAGAGAGCTTCCCTCT 889
 QY 946 TGTGCGCTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTTCCGGAATGTATGAT 1005
 Db 890 TGTGCGCTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTTCCGGAATGTATGAT 949
 QY 1006 CTTGGGATTTGGAATTTCTTGGCCCATCTTTCAGACCTGGGTGATATCAATCACCAGCA 1065
 Db 950 CTTGGGATTTGGAATTTCTTGGCCCATCTTTCAGACCTGGGTGATATCAATCACCAGCA 1009
 QY 1066 GGGCATGGTGAACCACTTTCACACAAATGCCAGGAGTGCCTTATGTCCTCCAGGC 1125
 Db 1010 GGGCATGGTGAACCACTTTCACACAAATGCCAGGAGTGCCTTATGTCCTCCAGGC 1069
 QY 1126 TTTGGCGCTGCTTCTTGGCCCATATGAAAGTTTCTGCACTTCAGAAAGCTTGTGCGATG 1185
 Db 1070 TTTGGCGCTGCTTCTTGGCCCATATGAAAGTTTCTGCACTTCAGAAAGCTTGTGCGATG 1129
 QY 1186 CATGAATTTCTGGGCTTACCCGCTTCACTGATCTGAACAGAGAGAAACCTGTTGCT 1245
 Db 1130 CATGAATTTCTGGGCTTACCCGCTTCACTGATCTGAACAGAGAGAAACCTGTTGCT 1189
 QY 1246 GGAATCTTCTGCTCACTGAGTCTGCTGAGCAATCACTAAGAGGTTGAGAGGCTT 1305
 Db 1190 GGAATCTTCTGCTCACTGAGTCTGCTGAGCAATCACTAAGAGGTTGAGAGGCTT 1249
 QY 1306 TGTGCGACCTGCTGTCAGGCTCAGTCACTTCTGTAATGCTTGTGAGCCCTTCA 1365
 Db 1250 TGTGCGACCTGCTGTCAGGCTCAGTCACTTCTGTAATGCTTGTGAGCCCTTCA 1309
 QY 1366 TCTCTGAGCTTACTACATGCTCTGTGGGTATCACTGATGATGTTGTCCACAG 1425
 Db 1310 TCTCTGAGCTTACTACATGCTCTGTGGGTATCACTGATGATGTTGTGTCCACAG 1369
 QY 1426 TGTCAAGCAGAGAGACTTTTGTGTCATGCTTGTCTGTAAGAAAGAGAGTGGGAACT 1485
 Db 1370 TGTCAAGCAGAGAGACTTTTGTGTCATGCTTGTCTGTAAGAAAGAGAGTGGGAACT 1429
 QY 1486 TATGTGAGCAGACATCCACCACTGAAGAGGGTATGCTCTTCTTCTTGTGATCT 1545
 Db 1430 TATGTGAGCAGACATCCACCACTGAAGAGGGTATGCTCTTCTTCTTGTGATCT 1489
 QY 1546 TCTGCTGTTGGGAGACTTTCAGAGACTTGTGGCTGAGAGGCTTATTAAGCAGCAGACT 1605
 Db 1490 TCTGCTGTTGGGAGACTTTCAGAGACTTGTGGCTGAGAGGCTTATTAAGCAGCAGACT 1549
 QY 1606 ATCAGTGAATTTGATCCATAAACTCCCTGTCCACATCTTGGCCAAATGGGGAATGATCT 1665
 Db 1550 ATCAGTGAATTTGATCCATAAACTCCCTGTCCACATCTTGGCCAAATGGGGAATGATCT 1609
 QY 1666 TTCAACCAAGAGCTCACACAGATTTTTCACAGAGATGC 1703
 Db 1610 TTCAACCAAGAGCTCACACAGATTTTTCACAGAGATGC 1647

RESULT 4
 AC010547

LOCUS AC010547 183228 bp DNA linear PRI 29-SEP-2001
 DEFINITION Homo sapiens chromosome 16 clone RP11-510M2, complete sequence.
 ACCESSION AC010547
 VERSION AC010547.9 GI:15808510
 KEYWORDS HTG.
 SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS DOE Joint Genome Institute.
 TITLE Sequencing of Human Chromosome 16
 JOURNAL Unpublished
 REFERENCE
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (29-SEP-2001) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 On Sep 29, 2001 this sequence version replaced gi:14589436.
 COMMENT
 Sequence Quality Assessment:
 This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 GenBank flat file format but are available as part
 of this entry's ASN.1 file.

 Location/Qualifiers
 source
 1. 183228
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-510M2"
 BASE COUNT 51901 a 42098 c 40026 g 49203 t
 ORIGIN
 Query Match 78.2%; Score 1597; DB 9; Length 183228;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 107 AAGTCTTCCACTTCAGCAGCATGCTACTGCTTAAAAAATGAAGCTCTGCTTCTG 166
 Db 142584 AAGTCTTCCACTTCAGCAGCATGCTACTGCTTAAAAAATGAAGCTCTGCTTCTG 142643
 QY 167 GTTTCACAGATGGGCACTTGGCTATTTCTTCCAGATGTACAGCAGCAACATCAGCTCC 226
 Db 142644 GTTTCACAGATGGGCACTTGGCTATTTCTTCCAGATGTACAGCAGCAACATCAGCTCC 142703
 QY 227 CTGTCTATGAAGGACAGCCGAGCGATGACGCTGCTGTCTTCTTCCGCGCTCT 286
 Db 142704 CTGTCTATGAAGGACAGCCGAGCGATGACGCTGCTGTCTTCTTCCGCGCTCT 142763
 QY 287 GGCCTCTCTTTTGTGGGGAGACTTTTGGGAGCAGCCAGATGTTTCTACCTGATGAG 346
 Db 142764 GGCCTCTCTTTTGTGGGGAGACTTTTGGGAGCAGCCAGATGTTTCTACCTGATGAG 142823

```

OY 347 CCCGCTGGCAGCTGTGATGACCTTCAAGCAGAGCAGCCGCTGATGTGCACATGGCT 406
    |||
    |||
    |||
Db 142824 CCCGCTGGCAGCTGTGATGACCTTCAAGCAGAGCAGCCGCTGATGTGCACATGGCT 142883
OY 407 GTGGGGGATCTGATACGCGGCTCTTCTGTGGATGATGAGCTCTTATAGCTTCAAG 466
    |||
    |||
    |||
Db 142884 GTGGGGGATCTGATACGCGGCTCTTCTGTGGATGATGAGCTCTTATAGCTTCAAG 142943
OY 467 GAACCTGTGTCCCGGAGACAGTCACAGCTCTTCACTGGAGAGACAGCCGCGCTGTGT 526
    |||
    |||
    |||
Db 142944 GAACCTGTGTCCCGGAGACAGTCACAGCTCTTCACTGGAGAGACAGCCGCGCTGTGT 143003
OY 527 TGTGACCTGTGTGTGATCAATCCCAAGATGAATATATCCCGGCGCTCACTGACAG 586
    |||
    |||
    |||
Db 143004 TGTGACCTGTGTGTGATCAATCCCAAGATGAATATATCCCGGCGCTCACTGACAG 143063
OY 587 GTTCCTGTGATCAACAGCCCTTGTGAGGTGGAGAGAGCCGCGCTCTTCACTGACAG 646
    |||
    |||
    |||
Db 143064 GTTCCTGTGATCAACAGCCCTTGTGAGGTGGAGAGAGCCGCGCTCTTCACTGACAG 143123
OY 647 GTGTGTCTCAAGAGAGTGGCTTCTCAACCTGCACTGCTACCTGCTGCTGAAAGAC 706
    |||
    |||
    |||
Db 143124 GTGTGTCTCAAGAGAGTGGCTTCTCAACCTGCACTGCTACCTGCTGCTGAAAGAC 143183
OY 707 CCTCTCCTCAACCTGCAATATGTCGACCTGTCCGGGACCCCGGCGCTGTCCGTTCC 766
    |||
    |||
    |||
Db 143184 CCTCTCCTCAACCTGCAATATGTCGACCTGTCCGGGACCCCGGCGCTGTCCGTTCC 143243
OY 767 CGAGAACGCAAAAGGAGATGTCATGATGATGATGATGATGATGATGATGATGATGATG 826
    |||
    |||
    |||
Db 143244 CGAGAACGCAAAAGGAGATGTCATGATGATGATGATGATGATGATGATGATGATGATG 143303
OY 827 CAAAACTCAAGAGAGAGACCAACCTCTACTATGATGATGATGATGATGATGATGATGATG 886
    |||
    |||
    |||
Db 143304 CAAAACTCAAGAGAGAGAGACCAACCTCTACTATGATGATGATGATGATGATGATGATGATG 143363
OY 887 CTGGAGATCTAACAAGACCTCAAGCTCTTCCCAAGCCCTGACAGAGAGCTTACTGCTT 946
    |||
    |||
    |||
Db 143364 CTGGAGATCTAACAAGACCTCAAGCTCTTCCCAAGCCCTGACAGAGAGAGCTTACTGCTT 143423
OY 947 GTGGCGTATGAGAGACCTGTGAGAGCCCTGTGGGCGCAGACCTCCGAGATGATGATGATG 1006
    |||
    |||
    |||
Db 143424 GTGGCGTATGAGAGACCTGTGAGAGCCCTGTGGGCGCAGACCTCCGAGATGATGATGATG 143483
OY 1007 GTGGGATTTGAAATTTCTTGGCCCATCTTCAAGACCTGGGTGCATTAACATCAACCCGAGGCAAG 1066
    |||
    |||
    |||
Db 143484 GTGGGATTTGAAATTTCTTGGCCCATCTTCAAGACCTGGGTGCATTAACATCAACCCGAGGCAAG 143543
OY 1067 GGCATGGGTGACAGCCTTCCACACAAATGCCAGGATGCCCTTAATGTCTTCCAGGCT 1126
    |||
    |||
    |||
Db 143544 GGCATGGGTGACAGCCTTCCACACAAATGCCAGGATGCCCTTAATGTCTTCCAGGCT 143603
OY 1127 TGGCGCTGTCTTGGCCATGATGAAAGGTTTCTGACCTCAGAAAGCTGTGGCGATGGC 1186
    |||
    |||
    |||
Db 143604 TGGCGCTGTCTTGGCCATGATGAAAGGTTTCTGACCTCAGAAAGCTGTGGCGATGGC 143663
OY 1187 ATGAATTTCTGGGCTACCGCAGCTGACATCTGAACAAGAACAGAAACCTGTTGCTG 1246
    |||
    |||
    |||
Db 143664 ATGAATTTCTGGGCTACCGCAGCTGACATCTGAACAAGAACAGAAACCTGTTGCTG 143723
OY 1247 GATCTTGTGTCTTACCTGAGACCTGTCCCTGTGAGCAATTCACATAAGAGGTTGAGAAAGCTTT 1306
    |||
    |||
    |||
Db 143724 GATCTTGTGTCTTACCTGAGACCTGTCCCTGTGAGCAATTCACATAAGAGGTTGAGAAAGCTTT 143783
OY 1307 GCTGCACCTGTGTGACAGCTGATCACTTCTGATGATGATGATGATGATGATGATGATGATG 1366
    |||
    |||
    |||
Db 143784 GCTGCACCTGTGTGACAGCTGATCACTTCTGATGATGATGATGATGATGATGATGATGATG 143843
OY 1367 GTCTGAGCCTTAACATGATGTGTGGGATATCAACATGATGATGATGATGATGATGATGATG 1426
    |||
    |||
    |||
Db 143844 GTCTGAGCCTTAACATGATGTGTGGGATATCAACATGATGATGATGATGATGATGATGATG 143903

```

```

OY 1427 GCTCAAGCAGAGAGACTTTTGTGTCCATGCTGTGTCTTGAAGAAACAGACTGGGAACTTT 1486
    |||
    |||
    |||
Db 143904 GCTCAAGCAGAGAGACTTTTGTGTCCATGCTGTGTCTTGAAGAAACAGACTGGGAACTTT 143963
OY 1487 ATGTGAGCAGCAGATCCACAGCTGTAAGAGGATATGCTCTTCTTCTTGTGATCTT 1546
    |||
    |||
    |||
Db 143964 ATGTGAGCAGCAGATCCACAGCTGTAAGAGGATATGCTCTTCTTCTTGTGATCTT 144023
OY 1547 CCTGTGTGGCAGACTTTCAGAGACTTGTGGCTGTGAGGCTTATTAAGCAGCAGACAGTA 1606
    |||
    |||
    |||
Db 144024 CCTGTGTGGCAGACTTTCAGAGACTTGTGGCTGTGAGGCTTATTAAGCAGCAGACAGTA 144083
OY 1607 TCAGTGAATTTGATTCATTAACCTCCCTGTCCACATCTTCCCAATGGGAATGATCTT 1666
    |||
    |||
    |||
Db 144084 TCAGTGAATTTGATTCATTAACCTCCCTGTCCACATCTTCCCAATGGGAATGATCTT 144143
OY 1667 TCACCAAGAGCTCACCAGCATTTTCCAGAGATGC 1703
    |||
    |||
    |||
Db 144144 TCACCAAGAGCTCACCAGCATTTTCCAGAGATGC 144180

```

```

RESULT 5
AC138848/c
LOCUS      206943 bp    DNA    linear    HTG 21-JAN-2003
DEFINITION Homo sapiens chromosome 5 clone RP11-1301B21, WORKING DRAFT
ACCESSION  AC138848
VERSION    AC138848.1 GI:27805260
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE  1 (bases 1 to 206943)
            DOE Joint Genome Institute.
            Unpublished
            2 (bases 1 to 206943)
            DOE Joint Genome Institute.
            Direct Submission
            Submitted (21-JAN-2003) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 2719263
            Center Clone name: RPCT-11_1301B21
            -----
            Summary Statistics
            Consensus quality: 202039 bases at least Q40
            Consensus quality: 202804 bases at least Q30
            Consensus quality: 203337 bases at least Q20
            Estimated insert size: 175000; agarose-IP estimation
            Estimated insert size: 206343; sum-of-coverage estimation
            Quality coverage: 15.28 in Q20 bases; agarose-IP estimation
            Quality coverage: 12.96 in Q20 bases; sum-of-coverage estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            1 1219: contig of 1219 bp in length
            * 1220 1319: gap of unknown length
            * 1320 2493: contig of 1174 bp in length
            * 2494 2593: gap of unknown length
            * 2594 3666: contig of 1073 bp in length
            * 3667 3766: gap of unknown length

```

```

* 3767 5167: contig of 1401 bp in length
* 5168 5267: gap of unknown length
* 5268 12818: contig of 7551 bp in length
* 12819 12918: gap of unknown length
* 12919 36169: contig of 23251 bp in length
* 36170 36269: gap of unknown length
* 36270 206943: contig of 170674 bp in length.
Location/Qualifiers
1. 206943
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-1301B21"
/clone_lib="RP11 human BAC library 11"
BASE COUNT 56312 a 44122 c 46163 g 59225 t 1121 others
ORIGIN
Query Match 78.2%; Score 1597; DB 2; Length 206943;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 107 AAGGCTCCACCTCAGACACATGCTACTGCTTAAAAAATGAAGCTCTGCTTTCTG 166
|
|
|
Db 125368 AAGGCTCCACCTCAGACACATGCTACTGCTTAAAAAATGAAGCTCTGCTTTCTG 125309
|
|
|
Qy 167 GTTCCAGATGGCATCTGGCTCTATCTTCCACATGTACAGACACATGATGATGCTCC 226
|
|
|
Db 125308 GTTCCAGATGGCATCTGGCTCTATCTTCCACATGTACAGACACATGATGATGCTCC 125249
|
|
|
Qy 227 CTGTATGAAGGACACAGCCGAGCGATGACAGCTGCTGTCTCTCTCTCTGCGCTCT 286
|
|
|
Db 125248 CTGTATGAAGGACACAGCCGAGCGATGACAGCTGCTGTCTCTCTCTCTGCGCTCT 125189
|
|
|
Qy 287 GGCCTCTCTTTTGGGGGACCTTTTGGGAGACACCAATGTTTCTTACTGATGAGAG 346
|
|
|
Db 125188 GGCCTCTCTTTTGGGGGACCTTTTGGGAGACACCAATGTTTCTTACTGATGAGAG 125129
|
|
|
Qy 347 CCCGCTGGACAGTGTGATGACCTTCAAGAGAGACACCGCTGGATGCTGACATGCT 406
|
|
|
Db 125128 CCCGCTGGACAGTGTGATGACCTTCAAGAGAGACACCGCTGGATGCTGACATGCT 125069
|
|
|
Qy 407 GTGCGGATCTGATACGGGCGCTTCTTGTGAGACATGACGCTTTGATGCTTACATG 466
|
|
|
Db 125068 GTGCGGATCTGATACGGGCGCTTCTTGTGAGACATGACGCTTTGATGCTTACATG 125009
|
|
|
Qy 467 GAACCTGGTCCCGGAGACAGTCCAGGCTTTTCAGTGGAGAAACGCGGCGCTGTGT 526
|
|
|
Db 125008 GAACCTGGTCCCGGAGACAGTCCAGGCTTTTCAGTGGAGAAACGCGGCGCTGTGT 124949
|
|
|
Qy 527 TCTGACCTGCTGTGATGATCCCAAGATGAATCATCCCGGCGCTGATGACAG 586
|
|
|
Db 124948 TCTGACCTGCTGTGATGATCCCAAGATGAATCATCCCGGCGCTGATGACAG 124889
|
|
|
Qy 587 CTCCTGTGAGTCAACAGCCCTTTGAGGTGTGAGAGAGGCTCCGCTCTTACAGCCAC 646
|
|
|
Db 124888 CTCCTGTGAGTCAACAGCCCTTTGAGGTGTGAGAGAGGCTCCGCTCTTACAGCCAC 124829
|
|
|
Qy 647 GTGCTGCTCAAGAGGTGCTTCTTCAACCTGACGTCCCTCAACCGGCTGTGAAGAC 706
|
|
|
Db 124828 GTGCTGCTCAAGAGGTGCTTCTTCAACCTGACGTCCCTCAACCGGCTGTGAAGAC 124769
|
|
|
Qy 707 CCCGCCCAACCGCATGTGTGACCTGCTCCGGGACCCCGGCGCTTCTCCGTTCC 766
|
|
|
Db 124768 CCCGCCCAACCGCATGTGTGACCTGCTCCGGGACCCCGGCGCTTCTCCGTTCC 124709
|
|
|
Qy 767 CGAAGAGCAAGAGAGATCTCATGATTGACAGTGCATTTGATGGGCGACATGAG 826
|
|
|
Db 124708 CGAAGAGCAAGAGAGATCTCATGATTGACAGTGCATTTGATGGGCGACATGAG 124649
|
|
|
Qy 827 CAAAACCTCAAGAGAGAGCAACCTTACTATGTGATGACAGTCTATCTGCCAAAGCCAG 886
|
|
|
Db 124648 CAAAACCTCAAGAGAGAGCAACCTTACTATGTGATGACAGTCTATCTGCCAAAGCCAG 124589

```

```

Qy 887 CTGAGATCTCAAGACATCCAGTCCCTTGGCCCAAGGCCCTGACGAACGCTACTGCTT 946
|
|
|
Db 124588 CTGAGATCTCAAGACATCCAGTCCCTTGGCCCAAGGCCCTGACGAACGCTACTGCTT 124529
|
|
|
Qy 947 GTGCGTATGAGAGACTGGCTGAGGCCCTGTGGCCAGACTTCCGATATGAAATTC 1006
|
|
|
Db 124528 GTGCGTATGAGAGACTGGCTGAGGCCCTGTGGCCAGACTTCCGATATGAAATTC 124469
|
|
|
Qy 1007 GTGGATTTGAAATCTTGGCCCATCTTCAAGCTTGGGTGCTTAACATCACCAGGCAAG 1066
|
|
|
Db 124468 GTGGATTTGAAATCTTGGCCCATCTTCAAGCTTGGGTGCTTAACATCACCAGGCAAG 124409
|
|
|
Qy 1067 GGCATGGGTGACACAGCTTTCACACAAATGCAAGGATGGCCCTTAATGTCTCCAGGCT 1126
|
|
|
Db 124408 GGCATGGGTGACACAGCTTTCACACAAATGCAAGGATGGCCCTTAATGTCTCCAGGCT 124349
|
|
|
Qy 1127 TGGCGTGTCTTGGCCCTATGAAAGGTTTCTGACTTCAGAAAGCCTGTGGGATGCC 1186
|
|
|
Db 124348 TGGCGTGTCTTGGCCCTATGAAAGGTTTCTGACTTCAGAAAGCCTGTGGGATGCC 124289
|
|
|
Qy 1187 ATGAATTTGCTGGGCTACCGGACGTCAGATCTGAACAAAGAGAAACCTGTGCTG 1246
|
|
|
Db 124288 ATGAATTTGCTGGGCTACCGGACGTCAGATCTGAACAAAGAGAAACCTGTGCTG 124229
|
|
|
Qy 1247 GATCTTGTCTTACCTGAGCTGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAAAGCTTT 1306
|
|
|
Db 124228 GATCTTGTCTTACCTGAGCTGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAAAGCTTT 124169
|
|
|
Qy 1307 GCTGACACCTGTGTCAAGCTCAAGTCACTTCTCTGAATGCTTGAAGCTTGTGCTACAT 1366
|
|
|
Db 124168 GCTGACACCTGTGTCAAGCTCAAGTCACTTCTCTGAATGCTTGAAGCTTGTGCTACAT 124109
|
|
|
Qy 1367 CTCTGAGCCTTAACATACATGCTGTGGGTATCACACTAGTGTGATGTTGTCCACAGCT 1426
|
|
|
Db 124108 CTCTGAGCCTTAACATACATGCTGTGGGTATCACACTAGTGTGATGTTGTCCACAGCT 124049
|
|
|
Qy 1427 GCTCAAGAGAGAGACTTTTGTGTCACAGCTTGTGTCTAGAAAACAGACTGGGAACTT 1486
|
|
|
Db 124048 GCTCAAGAGAGAGACTTTTGTGTCACAGCTTGTGTCTAGAAAACAGACTGGGAACTT 123989
|
|
|
Qy 1487 ATGTGAGCAGACATCCACAGTGAAGAGGATGATGCTCTTCTTCTTCTGATCTT 1546
|
|
|
Db 123988 ATGTGAGCAGACATCCACAGTGAAGAGGATGATGCTCTTCTTCTTCTGATCTT 123929
|
|
|
Qy 1547 CCTGTCTGGGAGACTTATAGAGACTTGTGGCTGTGAGAGCCTATTAAACAGACACATA 1606
|
|
|
Db 123928 CCTGTCTGGGAGACTTATAGAGACTTGTGGCTGTGAGAGCCTATTAAACAGACACATA 123869
|
|
|
Qy 1607 TCAGTGAATGATGATCAATAACCTCCCTGTCACATCTTGGCCATGGGGAATGATCTT 1666
|
|
|
Db 123868 TCAGTGAATGATGATCAATAACCTCCCTGTCACATCTTGGCCATGGGGAATGATCTT 123809
|
|
|
Qy 1667 TCACCAAGAGACTCACAGACTTTTCCACAGAGATGC 1703
|
|
|
Db 123808 TCACCAAGAGACTCACAGACTTTTCCACAGAGATGC 123772
|
|
|
RESULT 6
LOCUS BD127258 1979 bp DNA linear PAT 18-SEP-2002
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD127258.1 GI:23222203
VERSION BD127258.1 JP 2002017375-A/2689.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1979)
Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.

```

TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 200201375-A 2689 22-JAN-2002;
HELIOS RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 200201375-A/2689
PD 22-JAN-2002 JP 20020253172
PI 07-JUL-2000 JP 20020253172
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
PI ISHII,
PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUJI OTSUKI, HISASHI KOHA

PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/PC
10, C12P21/02, C12Q1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT CDS (69)..(1226).

FEATURES
source Location/Qualifiers
1..1979
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT 451 a 542 c 474 g 512 t
ORIGIN

Query Match 75.7%; Score 1547; DB 6; Length 1979;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; gaps 0;

QY 106 CAAGGCTTCCACTTACAGCAATGCTACTGCTTAATAAATGAACTGCTGTTTCT 165
Db 47 CAAGGCTTCCACTTACAGCAATGCTACTGCTTAATAAATGAACTGCTGTTTCT 106
QY 166 GGTTCCTCCAGATGAGCATCTTGCTATCTTCCATGATACAGGCACAAATCAGCTC 225
Db 107 GGTTCCTCCAGATGAGCATCTTGCTATCTTCCATGATACAGGCACAAATCAGCTC 166
QY 226 CCTGTCTAAGAAAGCAGACAGCCGAGCGATGACAGTGTCTGTCTTCTTCCGCGCTC 285
Db 167 CCTGTCTAAGAAAGCAGACAGCCGAGCGATGACAGTGTCTGTCTTCTTCCGCGCTC 226
QY 286 TGGCTCTCTTTTGTGGGAGAGCTTTTGGGAGAGCCAGAGTTTCTTACTGATGGA 345
Db 227 TGGCTCTCTTTTGTGGGAGAGCTTTTGGGAGAGCCAGAGTTTCTTACTGATGGA 286
QY 346 GCCCGCTGCGACAGTGTGATGACCTTCAAGCAGACAGCCGCTGATGCTGCACATGCG 405
Db 287 GCCCGCTGCGACAGTGTGATGACCTTCAAGCAGACAGCCGCTGATGCTGCACATGCG 346
QY 406 TGTGCGGATCTGATACAGGCGCTCTTGTGTCGACATGAGCGTCTTGATGCTACAT 465
Db 347 TGTGCGGATCTGATACAGGCGCTCTTGTGTCGACATGAGCGTCTTGATGCTACAT 406
QY 466 GGAACCTGATCCCGGAGACAGTCCAGCCCTTTTCAGTGGGAGAAAGCCGCGCTCTG 525
Db 407 GGAACCTGATCCCGGAGACAGTCCAGCCCTTTTCAGTGGGAGAAAGCCGCGCTCTG 466
QY 526 TTCTGCACTGCTGTGATCATCCCAAGATGAATTCATCCCGGCGTCTACTGAG 585
Db 467 TTCTGCACTGCTGTGATCATCCCAAGATGAATTCATCCCGGCGTCTACTGAG 526
QY 586 GCTCTGTGCAAGTCAAGAGCCCTTTGAGGTGTGAGAAAGCCGCTGCTCTACAGCA 645
Db 527 GCTCTGTGCAAGTCAAGAGCCCTTTGAGGTGTGAGAAAGCCGCTGCTCTACAGCA 586
QY 646 CGTGGGTCAAGAGAGTGGCTTCTTCAACCTGAGTCCCTTACCGCTGTGAAAGA 705
Db 587 CGTGGGTCAAGAGAGTGGCTTCTTCAACCTGAGTCCCTTACCGCTGTGAAAGA 646
QY 706 CCCCTCCCTCAACCTGATGCTGACCTGTGTCGGAGAGCCCGGCGGTGTCCGTTT 765
Db 706 CCCCTCCCTCAACCTGATGCTGACCTGTGTCGGAGAGCCCGGCGGTGTCCGTTT 765

Db 647 CCCCTCCCTCAACCTGATGCTGACCTGTGTCGGAGAGCCCGGCGGTGTCCGTTT 706
QY 766 CCGAGAGCGACAAAGAGAGATCTCATGATGACAGTGGCATTTGTATGGGAGCATGA 825
Db 707 CCGAGAGCGACAAAGAGAGATCTCATGATGACAGTGGCATTTGTATGGGAGCATGA 766
QY 826 GCAAAAATCTCAAGAGAGAGAGACCACTATGATGATGACAGTATGCTGCAAGCA 885
Db 767 GCAAAAATCTCAAGAGAGAGAGACCACTATGATGATGACAGTATGCTGCAAGCA 826
QY 886 GCTGAGATCTCAAAACCATCCAGTCTTGGCCCAAGGCCCTCCAGAAAGCTACTGCT 945
Db 827 GCTGAGATCTCAAAACCATCCAGTCTTGGCCCAAGGCCCTCCAGAAAGCTACTGCT 886
QY 946 TGTGCGCTATGAGAGACCTGGCTGAGCCCTGTGGCCCAAGACTTCCGATGATGAAT 1005
Db 887 TGTGCGCTATGAGAGACCTGGCTGAGCCCTGTGGCCCAAGACTTCCGATGATGAAT 946
QY 1006 CGTGGATTTGGAATTTCTTGGCCCATCTTCCAGACTTGGGTGATTAATCATCCGAGCA 1065
Db 947 CGTGGATTTGGAATTTCTTGGCCCATCTTCCAGACTTGGGTGATTAATCATCCGAGCA 1006
QY 1066 GGGCATGGGTGACAGGCTTTCCACAAATGCGAGGATGCCCTTAATGTCCGAGG 1125
Db 1007 GGGCATGGGTGACAGGCTTTCCACAAATGCGAGGATGCCCTTAATGTCCGAGG 1066
QY 1126 TTGGCGCTGCTTTTCCCTATGAAAAAGTTTCTCAGCTTCAGAAAAAGCTTGGCGATGC 1185
Db 1067 TTGGCGCTGCTTTTCCCTATGAAAAAGTTTCTCAGCTTCAGAAAAAGCTTGGCGATGC 1126
QY 1186 CATGAATTTGCTGGGTACCGCCAGCTCAGATTTGAACAAGAACAGAAAGCTGTTCCT 1245
Db 1127 CATGAATTTGCTGGGTACCGCCAGCTCAGATTTGAACAAGAACAGAAAGCTGTTCCT 1186
QY 1246 GGAATTTCTGTCTACCTGAGCTGCTCCCTGAGCAAAATCCACTAAGAGGTTGAGAAAGCTT 1305
Db 1187 GGAATTTCTGTCTACCTGAGCTGCTCCCTGAGCAAAATCCACTAAGAGGTTGAGAAAGCTT 1246
QY 1306 TGTGTCACCTGCTGTGACCTCAGTCACTTCTGTGAATGCTTGTGAGCTTGGCTTACA 1365
Db 1247 TGTGTCACCTGCTGTGACCTCAGTCACTTCTGTGAATGCTTGTGAGCTTGGCTTACA 1306
QY 1366 TGTGTCACCTGCTGTGACCTCAGTCACTTCTGTGAATGCTTGTGAGCTTGGCTTACA 1425
Db 1307 TGTGTCACCTGCTGTGACCTCAGTCACTTCTGTGAATGCTTGTGAGCTTGGCTTACA 1366
QY 1426 TGTGTCACCTGCTGTGACCTCAGTCACTTCTGTGAATGCTTGTGAGCTTGGCTTACA 1485
Db 1367 TGTGTCACCTGCTGTGACCTCAGTCACTTCTGTGAATGCTTGTGAGCTTGGCTTACA 1426
QY 1486 TATGTAGACAGCATCCACAGTGAAGAGGATTTGCTCTTCTTCTTCTTCTTCTTCT 1545
Db 1427 TATGTAGACAGCATCCACAGTGAAGAGGATTTGCTCTTCTTCTTCTTCTTCTTCT 1486
QY 1546 TCTGTGTGGGAGAGCTTTCAGAGACTTGTGGGCTGAGAGCCCTTAATGACAGACAGT 1605
Db 1487 TCTGTGTGGGAGAGCTTTCAGAGACTTGTGGGCTGAGAGCCCTTAATGACAGACAGT 1546
QY 1606 ATCAGTGAATTTGATCATTAACCTCCCTGTCACATCTTGGCCCAATGGGGAATGATCT 1665
Db 1547 ATCAGTGAATTTGATCATTAACCTCCCTGTCACATCTTGGCCCAATGGGGAATGATCT 1606
QY 1666 TTACCAAAAGAGCTACACAGACTTTTCCACAGAGATGC 1703
Db 1607 TTACCAAAAGAGCTACACAGACTTTTCCACAGAGATGC 1644

RESULT 7
AK074746 1979 bp mRNA linear PRI 03-SEP-2002
LOCUS AK074746 Homo sapiens cDNA FLJ90265 fis, clone NTBRM4001325, weakly similar
DEFINITION to CHONDROITIN 6-SULFOTRANSFERASE (EC 2.8.2.17).
ACCESSION AK074746

VERSION AK074746.1 GI:22760388
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Iisogai,T., Ota,T., Nishikawa,T., Hayashi,K., Otsuki,T.,
 Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S.,
 Kawai-Hio,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y.,
 Kojima,S., Nagahara,K., Masuhara,Y., Ono,T., Okano,K., Yoshikawa,Y.,
 Aotsuka,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and
 Nishimura,K.
 TITLE NEDO human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1979)
 AUTHORS Iisogai,T. and Otsuki,T.
 TITLE Direct Submission
 JOURNAL Submitted (25-MAR-2002) Takao Iisogai, Helix Research Institute,
 Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
 COMMENT NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:
 Research Association for Biotechnology; cDNA library construction:
 Institute of Medical Science, University of Tokyo; Laboratory of
 Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass
 sequencing and clone selection: Helix Research Institute (supported
 by Japan Key Technology Center etc.).
 FEATURES
 source
 1. 1979
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="NT2RM4001325"
 /cell_line="NT2"
 /cell_type="teratocarcinoma"
 /clone_id="NT2RM4"
 /note="cloning vector: PME18SFL3-mRNA from uninduced NT2
 neuronal precursor cells"
 69..1329
 /note="unamed protein product"
 /codon_start=1
 /protein_id="BAC11177.1"
 /db_xref="GI:22760389"
 /translation="MLPKKKKLLFLVSOMAILLFFPHMYSHNINSLSMKOPBRMH
 VIVLSMRSGSSFYGLPGCHPDVFLYMERPAHYHMTFKOSTAMHLHAYVDLIRAY
 LCDMSVDATHEPBPBROSLLFQENSRALCSAPACDILIPDEILIPRACHLDSQP
 FEVEKACRSTSHVLEKVERFNLQSLYPLIKDPSLNHIVLHVPDPAVERSRERT
 GDLMDIRIVWGQHEOKLKEKEDPYVYVQVQSOLEIKYKISLPKLOERYLYLVRY
 EDLARAVAOQTSRMTEFVGLFELPLQTVHNTIRGKMGMDHAFHTNARDALNTSQAM
 RMSLPYKVSRLKACGDAMNLLGVRHVRSEQBNLLDLSTWYPEQIH"
 BASE COUNT 451 a 542 c 474 g 512 t
 ORIGIN
 Query Match 75.7%; Score 1547; DB 9; Length 1979;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 346 GCCCGCTGGCAGCGTGTGATGACCTTCAAGCAGACACCGCGTGGATGTCACATGGC 405
 DB 287 GCCCGCTGGCAGCGTGTGATGACCTTCAAGCAGACACCGCGTGGATGTCACATGGC 346
 QY 406 TGTGGGAGATGATAGCGGCGCTCTTGTGGCAGATGACGCTTGTGATGCTACAT 465
 DB 347 TGTGGGAGATGATAGCGGCGCTCTTGTGGCAGATGACGCTTGTGATGCTACAT 406
 QY 466 GGAACCTGTGTCGGGAGACAGTCCAGCCTTTCAGTGGGAGAACACCGGCGCTGTG 525
 DB 407 GGAACCTGTGTCGGGAGACAGTCCAGCCTTTCAGTGGGAGAACACCGGCGCTGTG 466
 QY 526 TTCTGACCTGCTGTGACATATCCCAAGATGAATATATCCCGGCGCTGACATGAG 585
 DB 467 TTCTGACCTGCTGTGACATATCCCAAGATGAATATATCCCGGCGCTGACATGAG 526
 QY 586 GCTCTGTGCAATCAACAGCCCTTGTAGGTGTGAGAGAGGCGCTGCTCTACAGCA 645
 DB 527 GCTCTGTGCAATCAACAGCCCTTGTAGGTGTGAGAGAGGCGCTGCTCTACAGCA 586
 QY 646 CGTGTGCTCAAGAGGTGCGCTTCTTCAACCTGCAATCCCTTACCGGCTGTAAAGA 705
 DB 587 CGTGTGCTCAAGAGGTGCGCTTCTTCAACCTGCAATCCCTTACCGGCTGTAAAGA 646
 QY 706 CCCCTCCCTCAACCTGCAATGCTGACCTGCTGCGGAGCCCGGCGGCGTTCGTTG 765
 DB 647 CCCCTCCCTCAACCTGCAATGCTGACCTGCTGCGGAGCCCGGCGGCGTTCGTTG 706
 QY 766 CCGAGAGCAGCAAGAGGAGATCTCATGATGACAGTGTGATGTGATGGGACAGATGA 825
 DB 707 CCGAGAGCAGCAAGAGGAGATCTCATGATGACAGTGTGATGTGATGGGACAGATGA 766
 QY 826 GCAAAACTCAAGAGAGAGACCAACCTTATGTGATGACAGTGTGATGATGATGATG 885
 DB 767 GCAAAACTCAAGAGAGAGACCAACCTTATGTGATGACAGTGTGATGATGATGATG 826
 QY 886 GCTGAGATGACAGACATCATGCTGCTGCGGAGCGCGGAGAGAGGCTACCTGCT 945
 DB 827 GCTGAGATGACAGACATCATGCTGCTGCGGAGCGCGGAGAGAGGCTACCTGCT 886
 QY 946 TGTGGCGTATGAGACCTGTGCTGAGCCCTGTGGCCAGAGCTTCCGATGTATGAAT 1005
 DB 887 TGTGGCGTATGAGACCTGTGCTGAGCCCTGTGGCCAGAGCTTCCGATGTATGAAT 946
 QY 1006 CGTGGGATTTGAATTTCTTCCCATCTTCAAGACCTGGGTGATTAATCATCACCAGCA 1065
 DB 947 CGTGGGATTTGAATTTCTTCCCATCTTCAAGACCTGGGTGATTAATCATCACCAGCA 1006
 QY 1066 GGGCATGGGTGACACAGGCTTTCACACAAATGCCAGAGTGGCCTTAATGTCTCCAGGC 1125
 DB 1007 GGGCATGGGTGACACAGGCTTTCACACAAATGCCAGAGTGGCCTTAATGTCTCCAGGC 1066
 QY 1126 TTGGCGTGTGCTTTCCTATGAAAGCTTTCGACTTCAGAAAGCCTGTGGGATGC 1185
 DB 1067 TTGGCGTGTGCTTTCCTATGAAAGCTTTCGACTTCAGAAAGCCTGTGGGATGC 1126
 QY 1186 CATGAATTTGCTGGGCTACCGGCAGTCAATCTTAACAAGAGAGAGAAACCTTGTCT 1245
 DB 1127 CATGAATTTGCTGGGCTACCGGCAGTCAATCTTAACAAGAGAGAGAAACCTTGTCT 1186
 QY 1246 GGATCTTGTGCTACCTGAGAGTGTCCGAGCAATTCACATTAAGAGGTTGAGAGGCTT 1305
 DB 1187 GGATCTTGTGCTACCTGAGAGTGTCCGAGCAATTCACATTAAGAGGTTGAGAGGCTT 1246
 QY 1306 TGTCTGCACTGTGTGTACAGCTCACTCACTTCTCTGAATGCTTGTGAGCCTGTCA 1365
 DB 1247 TGTCTGCACTGTGTGTACAGCTCACTCACTTCTCTGAATGCTTGTGAGCCTGTCA 1306
 QY 1366 TCTCTGAGCCTTAATCAATGATGTGTGGGTATCAACAGAGTGTGTGTGTGTGTGTGT 1425
 DB 1307 TCTCTGAGCCTTAATCAATGATGTGTGGGTATCAACAGAGTGTGTGTGTGTGTGTGT 1366

Db	1540	TATGTGAGCAGCATCCACCACTGGAAGAAGGGTATTGTCCTTTCTTTTGATNCT	1592
OY	1546	TCCTGTCTGGGCACACTTTCAGAGACTTTGTGGCTCGAAGCCTTTTAAGCACGACAGT	1605
Db	1600	TTCTGTCTGGGCACACTTTCAGAGACTTTGTGGCTCGAAGCCTTTTAAGCACGACAGT	1659
OY	1606	ATCAGTGAATGATTCATAAACTCCCTGTCACATCTTCCCAATGGGAATGAGATCT	1665
Db	1660	ATCAGTGAATGATTCATAAACTCCCTGTCACATCTTCCCAATGGGAATGAGATCT	1719
OY	1666	TTTCACCAAAGAGCTCACCAGCATTTTTCACAGAGATGC	1703
Db	1720	TTTCACCAAAGAGCTCACCAGCATTTTTCACAGAGATGC	1757
RESULT 10			
LOCUS	AF131235	2032 bp	mRNA
DEFINITION	Homo sapiens N-acetylgucosamine 6-O-sulfotransferase mRNA,		linear PRI 10-AUG-1999
ACCESSION	AF131235		complete cds.
VERSION	AF131235.1		GI:4927113
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 2032)		
JOURNAL	Bistrup,A., Bhakta,S., Lee,J.K., Belov,Y.Y., Gunn,M.D., Zuo,F.R.,		
MEDLINE	Huang,C.C., Kannagi,R., Rosen,S.D. and Hemmerich,S.		
PUBMED	Sulfotransferases of two specificities function in the		
REFERENCE	reconstitution of high endothelial cell ligands for L-selectin		
AUTHORS	J. Cell Biol. 145 (4), 899-910 (1999)		
	99264336		
	10330415		
	2 (bases 1 to 2032)		
	Bistrup,A., Tangemann,K., Bhakta,S., Lee,J.-K., Belov,Y.Y.,		
	Gunn,M.D., Zuo,F.-R., Huang,C.-C., Kannagi,R., Rosen,S.D. and		
	Hemmerich,S.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-FEB-1999) Respiratory Diseases, Roche Bioscience,		
FEATURES	3401 Hillview Avenue, Palo Alto, CA 94304, USA		
SOURCE	Location/Qualifiers		
	1..2032		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/cell_type="high endothelial"		
	/tissue_type="tonsil"		
	182..1342		
CDS	/function="biosynthesis of endothelial ligands for		
	L-selectin."		
	/note="HEC-GlcNAc-6-ST"		
	/codon_start=1		
	/product="N-acetylglucosamine 6-O-sulfotransferase"		
	/protein_id="AAD33015.1"		
	/db_xref="GI:4927114"		
	/translation="MLPKMKLLFLVSNALILPFMYSHNISLSKAQPERHH		
	VIVLSWRSSSEPVGLFGFHDPVFYLMEPRAMHWTFKSTAMLMAYADLRAVLF		
	LCDMSVDAIEMEPEPROCSLPFOWENSRALCASAPCIDITODETIIPAHCKLICSDP		
	FVEVERCSRYSHVLYLKEVEFFNQSLSPYLKLPSNLIVHLVRDPAVERSRERTK		
	EDLAIRSIWVGQHEOKLKREDDPYVMVVICOSQLIKTIQSILPALEERYLLVRY		
	GDIAIRVAQTSRMTEFEVGLFELPHLOTWNHTIRGKGMDHAFTVARDALNVSAQM		
	RMSLPYKVSRLRKACGDAMNLIGLRVRSBOEQRNLLDLLSTWTVPDQH"		
BASE COUNT	468 a	569 c	490 g
ORIGIN			505 t
Query Match	75.7%	Score 1547;	DB 9; Length 2032;
Best Local Similarity	99.9%;	Prod. No. 0;	
Matches 1597; Conservative	0;	Mismatches	1; Indels
		Gaps	0;
OY	106	CAAGCTTTCACATTCAGACAACAAGCTACACGCCCTAAAAAATGAAGCTCCGCTGTTTCT	165

Db	160	AAAGTCTTCCACTTCACGACACAATGCTACTGCTATAAAAAATGAAAGCTCGTGTTTCT	219
OY	166	GGTTTCCAGATGGCCATCTTGCGCTATTTCTTCACATGTACAGCCCAACATCAGCTC	225
Db	220	GGTTTCCAGATGGCCATCTTGCGCTATTTCTTCACATGTACAGCCCAACATCAGCTC	279
OY	226	CCTGCTATGAAAGGACAGACCGGAGCGGATCAGATGCTGGTTCTGTTCTCTGGCGCTC	285
Db	280	CCTGCTATGAAAGGACAGACCGGAGCGGATCAGATGCTGGTTCTGTTCTCTGGCGCTC	339
OY	286	TGCGTCTTCTTTTGTGGGGCAGCTTTTGGGAGCACCAGATGTTTCTACCTGATGA	345
Db	340	TGCGTCTTCTTTTGTGGGGCAGCTTTTGGGAGCACCAGATGTTTCTACCTGATGA	399
OY	346	GCCCGCCCTGGGACCGTGTGATGACTTTAAAGAGAGCAACCGCTGGATGGCAGATAGC	405
Db	400	GCCCGCCCTGGGACCGTGTGATGACTTTAAAGAGAGCAACCGCTGGATGGCAGATAGC	459
OY	406	TGTGGGGATCTGATACGGGCCGTCCTTCTTGTGCACATGAGCGCTTTTGTATGCGCTACAT	465
Db	460	TGTGGGGATCTGATACGGGCCGTCCTTCTTGTGCACATGAGCGCTTTTGTATGCGCTACAT	519
OY	466	GGAACCTGTGTCCTGGGAGACAGTCCAGCTCTTTTCAGTGGGAGAAACACCGGGCCCTGTG	525
Db	520	GGAACCTGTGTCCTGGGAGACAGTCCAGCTCTTTTCAGTGGGAGAAACACCGGGCCCTGTG	579
OY	526	TTTCTGACCTGCTCTGTGACATCATCCCAAGATGAATTCATCCCGCGGCTCACTGAG	585
Db	580	TTTCTGACCTGCTCTGTGACATCATCCCAAGATGAATTCATCCCGCGGCTCACTGAG	639
OY	586	GCTCCTGTGACATCAACAGCCCTTTTGTAGGTGTGTGAGAGAGGCTCGCTCCTTACAGCCA	645
Db	640	GCTCCTGTGACATCAACAGCCCTTTTGTAGGTGTGTGAGAGAGGCTCGCTCCTTACAGCCA	699
OY	646	CGTGGTGTCGAAGGAGGTGGCCTTCTTAACCTGTAGTCCCTTACCGCGTGTGAAGA	705
Db	700	CGTGGTGTCGAAGGAGGTGGCCTTCTTAACCTGTAGTCCCTTACCGCGTGTGAAGA	759
OY	706	CCCTCCCTCAACCTGCATATCGTGACCTGGTCCGGGACCCCGGGCGGTGTTCCGTTTC	765
Db	760	CCCTCCCTCAACCTGCATATCGTGACCTGGTCCGGGACCCCGGGCGGTGTTCCGTTTC	819
OY	766	CCGGAACGGCAAAAGGAGATTCATGATGTAGACAGTCCGCAATTGTGATGGGGGACAGTGA	825
Db	820	CCGGAACGGCAAAAGGAGATTCATGATGTAGACAGTCCGCAATTGTGATGGGGGACAGTGA	879
OY	826	GCAAAAACCTCAAGAAGAGAGCAACAACCTTACTATGTGATGCAAGGTCACTGCCAAAGCCA	885
Db	880	GCAGAAATCTCAAGAAGAGAGCAACAACCTTACTATGTGATGCAAGGTCACTGCCAAAGCCA	939
OY	886	GCTGGAGATCTACAAAGNCAATCCAGTCTTGTGCCAAGGCCCTGCAGAGAAAGCTTACCTGCT	945
Db	940	GCTGGAGATCTACAAAGNCAATCCAGTCTTGTGCCAAGGCCCTGCAGAGAAAGCTTACCTGCT	999
OY	946	TGTGGCTATAGGACCTGCGCTCGAGAGCCCTGTGGGCCCAAGATTCGCCAATGTATGAATT	1005
Db	1000	TGTGGCTATAGGACCTGCGCTCGAGAGCCCTGTGGGCCCAAGATTCGCCAATGTATGAATT	1059
OY	1006	CGTGGGATTTGGAATTTCTTGCCCATCTTCAAGACTGGGTGCATACATCACCCGAGGCCA	1065
Db	1060	CGTGGGATTTGGAATTTCTTGCCCATCTTCAAGACTGGGTGCATACATCACCCGAGGCCA	1119
OY	1066	GGGATGGGATACCAACGCTTTCACACAAATGSCAGGGATCCCTTAATGTCTCCAGGC	1125
Db	1120	GGGATGGGATACCAACGCTTTCACACAAATGSCAGGGATCCCTTAATGTCTCCAGGC	1179
OY	1126	TTGGCGCTGTCTTTGCCCTATGAAAAGTTTCTGCATTCAGAAAGCCTGTGGCGATGC	1185
Db	1180	TTGGCGCTGTCTTTGCCCTATGAAAAGTTTCTGCATTCAGAAAGCCTGTGGCGATGC	1239
OY	1186	CATGAATTTGTGGGCTTACCGGCCACGTACAGTCTGAAACAACAAGAAACCTGTGCTCT	1245
Db	1240	CATGAATTTGTGGGCTTACCGGCCACGTACAGTCTGAAACAACAAGAAACCTGTGCTCT	1299

QY 1067 GGCATGGGTGACACGCTTTCCACACAAATGCCAGGATGCCCTTAATGTCCTCCAGGCT 1126
|||||
Db 1032 GGCATGGGTGACACGCTTTCCACACAAATGCCAGGATGCCCTTAATGTCCTCCAGGCT 1091
QY 1127 TGGCGCTGGTCTTTGCCCTATGAAAAGTTTCTCCAGCTTCAGAAACCTGTGGCATGCC 1186
Db 1092 TGGCGCTGGTCTTTGCCCTATGAAAAGTTTCTCCAGCTTCAGAAACCTGTGGCATGCC 1151
QY 1187 ATGAATTTCTGGGGCTACCGCCACGTCAGATCTGAACAAAGAACAGAAACCGTGTGCTG 1246
Db 1152 ATGAATTTCTGGGGCTACCGCCACGTCAGATCTGAACAAAGAACAGAAACCGTGTGCTG 1211
QY 1247 GATCTTCTGTCTACCTGAGCTGTCCCTGAGCAATCCACTAAGAGGCTTGAGAGGCTTT 1306
Db 1212 GATCTTCTGTCTACCTGAGCTGTCCCTGAGCAATCCACTAAGAGGCTTGAGAGGCTTT 1271
QY 1307 GCTGCCACCTGGTGTACAGCTTACATGCTTCTGTAATGCTTCTGAGCTTGGCTTACAT 1366
Db 1272 GCTGCCACCTGGTGTACAGCTTACATGCTTCTGTAATGCTTCTGAGCTTGGCTTACAT 1331
QY 1367 CT 1368
Db 1332 CT 1333

RESULT 12
AR310485 2065 bp DNA linear PAT 12-JUN-2003
LOCUS AR310485 Sequence 26 from patent US 6558935.
DEFINITION AR310485
ACCESSION AR310485
VERSION AR310485.1 GI:31703448
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2065)
AUTHORS Tang,Y.T., Corley,N.C., Guegler,K.J., Baughn,M.R., Lal,P., Yue,H.,
Hillman,J.L. and Azimzai,Y.
TITLE Human transferase proteins
JOURNAL Patent: US 6558935-A 26 06-MAY-2003;
FEATURES
source 1. 2065
location/Qualifiers
BASE COUNT 466 a 575 c 491 g 533 t
ORIGIN

Query Match 56.4%; Score 1153; DB 6; Length 2065;
Best local similarity 99.6%; Pred. No. 0;
Matches 1593; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 106 CAAGCTCTTCACATTCAGCACAATGCTACTGCTAATAAAATGAAGCTCTGCTTTCT 165
Db 152 CAAGCTCTTCACATTCAGCACAATGCTACTGCTAATAAAATGAAGCTCTGCTTTCT 211
QY 166 GGTTCCTCCAGATGGGCATGTTGGCTATCTTCCATGATACAGCACAATCAGCTC 225
Db 212 GGTTCCTCCAGATGGGCATGTTGGCTATCTTCCATGATACAGCACAATCAGCTC 271
QY 226 CCTGTCTATGAAGCACAACCCGAGGAGCATGACGCTGTCTCTCTCCGAGGCTC 285
Db 272 CCTGTCTATGAAGCACAACCCGAGGAGCATGACGCTGTCTCTCTCCGAGGCTC 331
QY 286 TGGCTCTCTTTTGGGGGAGCTTTTGGGAGCACCAGATGTTTCTTACCTGATGA 345
Db 332 TGGCTCTCTTTTGGGGGAGCTTTTGGGAGCACCAGATGTTTCTTACCTGATGA 391
QY 346 GCGCGCTGGCAGCAGTGTGATGACCTTCAAGAGAGACCCGCTGATCTGACATGGC 405
Db 392 GCGCGCTGGCAGCAGTGTGATGACCTTCAAGAGAGACCCGCTGATCTGACATGGC 451
QY 406 TGTGGGAGATCTGATACGGGCGCTCTTGTGCGCATGAGCGTCTTGTGATGCTTACAT 465
|||||

Db 452 TGTGCGGATCTGATACGGGCGCTCTTGTGCGCATGAGCGTCTTGTATGCTTACAT 511
QY 466 GGAACCTGGTCCCGGAGACAGTCCAGCCCTTTCAGTGGAGAACAGCCGGGCGCTGTG 525
Db 512 GGAACCTGGTCCCGGAGACAGTCCAGCCCTTTCAGTGGAGAACAGCCGGGCGCTGTG 571
QY 526 TTTCTGACCTGGCTGTGACATCATCCACAGATGAATCATCCCGGGGCTCATGAC 585
Db 572 TTTCTGACCTGGCTGTGACATCATCCACAGATGAATCATCCCGGGGCTCATGAC 630
QY 586 GCTCCTGTGACATCAACACCCCTTTG-AGTGTGGAGAAAGGCTCCCTCTACAGCC 644
Db 631 GCTCCTGTGACATCAACACCCCTTTGAAATTTGTGGAAGAAAGCTCCCTCTACAGCC 690
QY 645 AGTGTGTCTCAAGAGAGTGGCGTCTTCAACCTGAGTCCCTCTACCGCTGCTGAAG 704
Db 691 AGTGTGTCTCAAGAGAGTGGCGTCTTCAACCTGAGTCCCTCTACCGCTGCTGAAG 750
QY 705 ACCCTCTCCTCAACCTGATATGCTGACACCTGGTCCGGAGCCCGGGCGGTTCGTT 764
Db 751 ACCCTCTCCTCAACCTGATATGCTGACACCTGGTCCGGAGCCCGGGCGGTTCGTT 810
QY 765 CCGGAGAACGCAAAAGGAGATCTCATGATGACAGTGGCATTTGTATGGGGGACAGAG 824
Db 811 CCGGAGAACGCAAAAGGAGATCTCATGATGACAGTGGCATTTGTATGGGGGACAGAG 870
QY 825 ACCAAATCTCAAGAGAGAGACCAACCTCTACTATGTATGAGAGTCTGCAAGCC 884
Db 871 AGCAGAAATCTCAAGAGAGAGACCAACCTCTACTATGTATGAGAGTCTGCAAGCC 930
QY 885 AGCTGAGATCTTCAAGAGACCATTCAGTCTTGGCCAGAGCCCTGCAAGAAAGCTTCTGC 944
Db 931 AGCTGAGATCTTCAAGAGACCATTCAGTCTTGGCCAGAGCCCTGCAAGAAAGCTTCTGC 990
QY 945 TTGTGCGATGAGAGCTGGTCCGAGCCCTGTGGGCCAGACTTCCGAATGATGAAT 1004
Db 991 TTGTGCGATGAGAGCTGGTCCGAGCCCTGTGGGCCAGACTTCCGAATGATGAAT 1050
QY 1005 TGTGTGATTTGGAATTTTGGCCCATCTTTCAGACCTGGTGTGATTAACATCACCCGAGCA 1064
Db 1051 TGTGTGATTTGGAATTTTGGCCCATCTTTCAGACCTGGTGTGATTAACATCACCCGAGCA 1110
QY 1065 AAGGCAATGGGTGACCAACGCTTTCCACACAAAGCCAGGATGCCCTTATGTCCTCCAG 1124
Db 1111 AAGGCAATGGGTGACCAACGCTTTCCACACAAAGCCAGGATGCCCTTATGTCCTCCAG 1170
QY 1125 CTGGCGCTGTCTTGGCCATATGAAGATTTCTGCAATGAGAAAGCCTGTGGCATG 1184
Db 1171 CTGGCGCTGTCTTGGCCATATGAAGATTTCTGCAATGAGAAAGCCTGTGGCATG 1230
QY 1185 CCATGAATTTGCTGGGCTACCGCCACGTCAGATCTGTAACAGAACAGAAACCTGTGTC 1244
Db 1231 CCATGAATTTGCTGGGCTACCGCCACGTCAGATCTGTAACAGAACAGAAACCTGTGTC 1290
QY 1245 TGGATCTCTGTCTACCTGAGCTGTCCGAGCAATTCACATGAAGGGTTAGAAAGCT 1304
Db 1291 TGGATCTCTGTCTACCTGAGCTGTCCGAGCAATTCACATGAAGGGTTAGAAAGCT 1350
QY 1305 TTGCTGCAACAGTGTGTCAGCTCAGCTCACTTCTCTGTAATCTTCTGAGCTTGGCTTAC 1364
Db 1351 TTGCTGCAACAGTGTGTCAGCTCAGCTCAGCTCACTTCTCTGTAATCTTCTGAGCTTGGCTTAC 1410
QY 1365 ATCTGTAGACCTTAACATCATGTCTGTGGGTATCACACTGATGTGATGTGTGCCACAC 1424
Db 1411 ATCTGTAGACCTTAACATCATGTCTGTGGGTATCACACTGATGTGATGTGTGCCACAC 1470
QY 1425 GTGCTTAACAGAAAGACTTTTGTGTCATGCTGTGTGTAAAGAAACACATCGGGGAACC 1484
Db 1471 GTGCTTAACAGAAAGACTTTTGTGTCATGCTGTGTGTAAAGAAACACATCGGGGAACC 1530
QY 1485 TTATGTGACAGACATCCACACAGTGAAGAAAGGATTTGCTCTTCTTCTTCTTCTGATC 1544
Db 1531 TTATGTGACAGACATCCACACAGTGAAGAAAGGATTTGCTCTTCTTCTTCTTCTGATC 1590
|||||

```

QY 1545 TTCTGTCGTGGGAGAGACTTTCAGAGACTTTGTGGCTGGAGCCCTATTAAAGCAGACAG 1604
DB 1591 TTCTGTCGTGGGAGAGACTTTCAGAGACTTTGTGGCTGGAGCCCTATTAAAGCAGACAG 1650
QY 1605 TATCAGTGAATGATTCATTAACCTCCCTGTCCACATCTTGCCCAATGGGAATGATC 1664
DB 1651 TATCAGTGAATGATTCATTAACCTCCCTGTCCACATCTTGCCCAATGGGAATGATC 1710
QY 1665 TTTCACCAAGAGCTCCAGCATTTTCCACAGAGATGC 1703
DB 1711 TTTCACCAAGAGCTCCAGCATTTTCCACAGAGATGC 1749

RESULT 13
LOCUS BD124832 877 bp DNA linear PAT 18-SEP-2002
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD124832
VERSION BD124832.1 GI:23219777
KEYWORDS JP 2002017375-A/263.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.
REFERENCE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002017375-A 263 22-JAN-2002;
HELIIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002017375-A/263
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253172
PI TOSHIO OTA,TERSTUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
PI ISHII,
PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TERSTUO OTSUKI,HISASHI KOGA
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
10,
PC C12P21/02,C12O1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH key
FEATURES
source
location/Qualifiers
FT source 1..877
/organism="Homo sapiens (human)"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT 183 a 266 c 221 g 203 t 4 others

ORIGIN
Query Match 34.0%; Score 695; DB 6; Length 877;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 CAAGGCTTCACATTCAGACAAATGCTACTGCTAAATAAATGAAGCTGCTGCTTCT 165
DB 47 CAAGGCTTCACATTCAGACAAATGCTACTGCTAAATAAATGAAGCTGCTGCTTCT 106
QY 166 GGTTCCTCCAGATGGCATCTTGCTATTTCTTCACATGATACAGCCACAATCAGCTC 225
DB 107 GGTTCCTCCAGATGGCATCTTGCTATTTCTTCACATGATACAGCCACAATCAGCTC 166
QY 226 CCGTGTATGAAGGCAAGCCGAGCCGATGACAGCTGCTGCTTCTGCTTCTGCGCTC 285
DB 167 CCGTGTATGAAGGCAAGCCGAGCCGATGACAGCTGCTGCTTCTGCTTCTGCGCTC 226

FEATURES
source
location/Qualifiers
FT source 1..877
/organism="Homo sapiens (human)"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

```

```

QY 286 TGCTCTCTTTTGTGGGAGAGCTTTTGGGACACACACAGATGTTTCTACCTGATGA 345
DB 227 TGCTCTCTTTTGTGGGAGAGCTTTTGGGACACACACAGATGTTTCTACCTGATGA 286
QY 346 GCCCGCTGGCAGCTGTGATGATCCTTCAACAGAGACACCGCTGATGCTGACATGCC 405
DB 287 GCCCGCTGGCAGCTGTGATGATCCTTCAACAGAGACACCGCTGATGCTGACATGCC 346
QY 406 TGTGGGAGATCTGATGATGAGGCGGCTCTTGTGGGACATGAGGCTCTTGTATGCTCAT 465
DB 347 TGTGGGAGATCTGATGATGAGGCGGCTCTTGTGGGACATGAGGCTCTTGTATGCTCAT 406
QY 466 GGAACCTGTGCTCCCGGAGACAGTCCAGCTCTTCACTGAGGAGAAACCGCGGCTGTG 525
DB 407 GGAACCTGTGCTCCCGGAGAGAGTCCAGCTCTTCACTGAGGAGAAACCGCGGCTGTG 466
QY 526 TTTCGACACTGCTGCTGATCATCCACAAGATGAATATCCCGGCTGCTGCTGACG 585
DB 467 TTTCGACACTGCTGCTGATCATCCACAAGATGAATATCCCGGCTGCTGCTGACG 526
QY 586 GCTCCTGTGACATCAAGAGCCCTTGAAGTGTGGAGAAAGGCGGCTGCTGACAGCA 645
DB 527 GCTCCTGTGACATCAAGAGCCCTTGAAGTGTGGAGAAAGGCGGCTGCTGACAGCA 586
QY 646 CGTGTGCTCAAGAGAGTGCCTTCTTCAACCTGACAGTCCCTTACCCGCTGTAAGA 705
DB 587 CGTGTGCTCAAGAGAGTGCCTTCTTCAACCTGACAGTCCCTTACCCGCTGTAAGA 646
QY 706 CCCCTCCTCAACCTGATATGATGACCTGCTGCGGAGCCCGGCGCTGCTGCTG 765
DB 647 CCCCTCCTCAACCTGATATGATGACCTGCTGCGGAGCCCGGCGCTGCTGCTG 706
QY 766 CCGAAGACGCAAGAGGAGATCTCATGATGATGA 800
DB 707 CCGAAGACGCAAGAGGAGATCTCATGATGATGA 741

RESULT 14
LOCUS BD126950 877 bp DNA linear PAT 18-SEP-2002
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD126950
VERSION BD126950.1 GI:2321895
KEYWORDS JP 2002017375-A/2381.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.
REFERENCE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002017375-A 2381 22-JAN-2002;
HELIIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002017375-A/2381
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253172
PI TOSHIO OTA,TERSTUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
PI ISHII,
PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TERSTUO OTSUKI,HISASHI KOGA
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
10,
PC C12P21/02,C12O1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH key
FEATURES
source
location/Qualifiers
FT source 1..877
/organism="Homo sapiens (human)"

```

```

source
1. .877
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT      183 a      266 c      221 g      203 t      4 others
ORIGIN

Query Match      34.0%; Score 695; DB 6; Length 877;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 CAGGCTTCCACTTCAGACAATGCTACTGCTTAATAAATAAAGAACTCTGCTGTTCT 165
    |||||||
DB 47 CAAAGCTTCCACTTCAGACAATGCTACTGCTTAATAAATAAAGAACTCTGCTGTTCT 106
    |||||||
QY 166 GATTTCACAGATGAGCATCTGCTATCTTCCACATGATACAGCACAACATCAGCTC 225
    |||||||
DB 107 GATTTCACAGATGAGCATCTGCTATCTTCCACATGATACAGCACAACATCAGCTC 166
    |||||||
QY 226 CCTGTCTAAGAAAGCAGACCCGAGGCGATGCAAGCTGTGTTCTGTCTGCGGCTC 285
    |||||||
DB 167 CCTGTCTAAGAAAGCAGACCCGAGGCGATGCAAGCTGTGTTCTGTCTGCGGCTC 226
    |||||||
QY 286 TGACCTCTCTTTTGTGTGGGAGAGTTTGTGGGAGACCCAGAGTTTCTACTGATGGA 345
    |||||||
DB 227 TGACCTCTCTTTTGTGTGGGAGAGTTTGTGGGAGACCCAGAGTTTCTACTGATGGA 286
    |||||||
QY 346 GCCCGGCTTGACAGCTGTGATGACCTTCAAGCAGAGACCGGCTGTGATGCTGACATGGC 405
    |||||||
DB 287 GCCCGGCTTGACAGCTGTGATGACCTTCAAGCAGAGACCGGCTGTGATGCTGACATGGC 346
    |||||||
QY 406 TGTGGGGATCTGATACGGGGCGCTCTTGTGTGGGAGATGAGGCTTGTGATGCTTACAT 465
    |||||||
DB 347 TGTGGGGATCTGATACGGGGCGCTCTTGTGTGGGAGATGAGGCTTGTGATGCTTACAT 406
    |||||||
QY 466 GGAACCTGTGCTCCCGAGAGAGTCCAGCCTTTTCAAGTGGGAGAAAGCGGGGCTGTG 525
    |||||||
DB 407 GGAACCTGTGCTCCCGAGAGAGTCCAGCCTTTTCAAGTGGGAGAAAGCGGGGCTGTG 466
    |||||||
QY 526 TTCTGACCTGCTGTGATCATCCCAAGATGAATCATCCCGGGGCTGCTGACAG 585
    |||||||
DB 467 TTCTGACCTGCTGTGATCATCCCAAGATGAATCATCCCGGGGCTGCTGACAG 526
    |||||||
QY 586 GGTCCGTGACAGCAACGCCCTTTGAGTGGTGGGAGAGGCTGCGGCTCTACAGCCA 645
    |||||||
DB 527 GGTCCGTGACAGCAACGCCCTTTGAGTGGTGGGAGAGGCTGCGGCTCTACAGCCA 586
    |||||||
QY 646 CGTGTGTCTAAGAGAGTCCGCTTCTTCAACTGACAGTCCCTTACCCCTGCTGAAGA 705
    |||||||
DB 587 CGTGTGTCTAAGAGAGTCCGCTTCTTCAACTGACAGTCCCTTACCCCTGCTGAAGA 646
    |||||||
QY 706 CCCCTTCCTCAACCTGATATGATGACCTGTGCGGAGCCCGGGGCTGCTGCTTC 765
    |||||||
DB 647 CCCCTTCCTCAACCTGATATGATGACCTGTGCGGAGCCCGGGGCTGCTGCTTC 706
    |||||||
QY 766 CCGAGAAGCACAAGAGGATCTCATGATTGACA 800
    |||||||
DB 707 CCGAGAAGCACAAGAGGATCTCATGATTGACA 741
    |||||||

RESULT 15
AX381256      517 bp      DNA      linear      PAT 18-MAR-2002
LOCUS      AX381256
DEFINITION      Sequence 194 from Patent WO0212280.
ACCESSION      AX381256
VERSION      AX381256.1 GI:19576075
KEYWORDS
SOURCE
ORGANISM      Homo sapiens (human)
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1      Pyle,R.A., Xu,J. and Secrist,H.

```

```

TITLE      Compositions and methods for the therapy and diagnosis of colon
            cancer
JOURNAL      Patent: WO 0212280-A 194 14-FEB-2002;
            CORIXA CORPORATION (US)
FEATURES
source
1. .517
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT      121 a      130 c      124 g      142 t
ORIGIN

Query Match      19.7%; Score 403; DB 6; Length 517;
Best Local Similarity 100.0%; Pred. No. 3.5e-228;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GGCTTGGCCAGCCGAGGTGATGAGCTGATGCTTCTGTAATGCTTGAAGCTTGC 1360
    |||||||
DB 13 GGCTTGGCCAGCCGAGGTGATGAGCTGATGCTTCTGTAATGCTTGAAGCTTGC 72
    |||||||
QY 1361 CTACATCTGAGCCTTAACATGATGCTGTGGGTATCACAGTGAAGTGTGATGCTC 1420
    |||||||
DB 73 CTACATCTGAGCCTTAACATGATGCTGTGGGTATCACAGTGAAGTGTGATGCTC 132
    |||||||
QY 1421 ACACGTGCTCAGACAGAGACTTTGTGTCCATGCTTGTGTCTAGAAAACAGACTGGGG 1480
    |||||||
DB 133 ACACGTGCTCAGACAGAGACTTTGTGTCCATGCTTGTGTCTAGAAAACAGACTGGGG 192
    |||||||
QY 1481 AACCTTATGAGACAGACATCCACAGTGAAGAGGTAAATGCTTCTTCTTCTT 1540
    |||||||
DB 193 AACCTTATGAGACAGACATCCACAGTGAAGAGGTAAATGCTTCTTCTTCTT 252
    |||||||
QY 1541 GATCTTCTGCTGTGGGAGACTTGAAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1600
    |||||||
DB 253 GATCTTCTGCTGTGGGAGACTTGAAGACTTGTGGCTGAGGCTTATTAAGCAGAC 312
    |||||||
QY 1601 ACAGTATCAGTGAATTAATGATTAACCTCCCTGTCACATCTTGGCCCAAGGGGAATG 1660
    |||||||
DB 313 ACAGTATCAGTGAATTAATGATTAACCTCCCTGTCACATCTTGGCCCAAGGGGAATG 372
    |||||||
QY 1661 GATCTTTCACAAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
    |||||||
DB 373 GATCTTTCACAAAGAGCTCACAGCATTTTCCACAGAGATGC 415
    |||||||

Search completed: August 14, 2003, 20:11:36
Job time : 7288 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 11:33:50 ; Search time 493.12 Seconds
(without alignments)
9271.801 Million cell updates/sec

Title: US-09-645-078-1
Perfect score: 2043
Sequence: 1 gaattcatgtgttggtta.....ttggatcctaataaaaaaa 2043

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1504479 seqs, 1118970152 residues

Total number of hits satisfying chosen parameters: 3008958

Initial DB seq length: 0
aximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCr_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCrUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1784.8	87.4	2032	9	US-09-816-825-1
2	1784.8	87.4	2032	13	US-10-007-262-2
3	480.2	23.5	517	10	US-09-819-580-194
4	478.8	23.4	505	10	US-09-998-598-2595
5	412.8	20.2	2544	9	US-09-927-602-1
6	336.2	16.5	48436	9	US-09-927-602-1
7	294.4	14.4	389	10	US-09-919-580-264
8	119.4	5.8	2393	9	US-09-833-790-244
9	119.4	5.8	2409	14	US-10-212-933-3
10	108	5.3	2150	14	US-10-212-933-1
11	62.6	3.1	731	9	US-09-735-705-63
12	62.6	3.1	731	10	US-09-850-716A-63
13	62.6	3.1	731	10	US-09-897-718-63
14	62.6	3.1	731	11	US-09-466-396A-63
15	62.6	3.1	731	12	US-10-117-982-63
16	59	2.9	303	9	US-09-833-790-243

17	59	2.9	303	9	US-09-833-790-401	Sequence 401, App
c 18	42.8	2.1	395	9	US-09-833-790-207	Sequence 207, App
c 19	42.4	2.1	2976	14	US-10-128-714-355	Sequence 355, App
c 20	42.4	2.1	2977	14	US-10-128-714-5355	Sequence 5355, App
21	41.6	2.0	720	13	US-10-027-632-22081	Sequence 22081, A
22	41.6	2.0	720	13	US-10-027-632-149086	Sequence 149086, A
23	41.4	2.0	1515	14	US-10-156-761-7394	Sequence 7394, App
24	41.4	2.0	9025608	14	US-10-156-761-1	Sequence 1, App1
25	40.2	2.0	1011	14	US-10-156-761-7058	Sequence 7058, App
c 26	40.2	2.0	9025608	14	US-10-156-761-1	Sequence 1, App1
27	40	2.0	2838	9	US-09-815-242-7888	Sequence 7888, App
28	39.6	1.9	344	9	US-09-864-761-19694	Sequence 19694, A
29	39.6	1.9	399	9	US-09-864-761-2913	Sequence 2913, App
30	39.6	1.9	738	14	US-10-156-761-1406	Sequence 1406, App
31	39.6	1.9	7092	14	US-10-037-270-252	Sequence 252, App
c 32	38.6	1.9	7269	14	US-10-037-270-251	Sequence 251, App
c 33	39.6	1.9	20165	13	US-10-159-330-7	Sequence 7, App1
c 34	39.6	1.9	20165	14	US-10-159-334-7	Sequence 7, App1
c 35	39.2	1.9	771	14	US-10-156-761-3385	Sequence 3385, App
36	38.2	1.9	1512	14	US-10-156-761-7404	Sequence 7404, App
c 37	38	1.9	666	13	US-10-027-632-222628	Sequence 222628, App
c 38	38	1.9	6381	12	US-10-311-455-939	Sequence 939, App
39	38	1.9	6381	12	US-10-240-485-71	Sequence 71, App1
40	37.8	1.9	963	14	US-10-156-761-5861	Sequence 5861, App
41	37.6	1.8	362	13	US-10-027-632-272279	Sequence 272279, App
c 42	37.6	1.8	1022	13	US-10-027-632-9247	Sequence 9247, App
c 43	37.6	1.8	1022	13	US-10-027-632-9248	Sequence 9248, App
c 44	37.6	1.8	6130	12	US-10-311-455-102	Sequence 102, App
c 45	37.6	1.8	7607	12	US-10-240-485-168	Sequence 168, App

ALIGNMENTS

RESULT 1	US-09-816-825-1	US-09-816-825-1
Sequence 1, Application US/09816825		
Patent No. US20010051370A1		
GENERAL INFORMATION:		
APPLICANT: Bistrup, Annette		
APPLICANT: Rosen, Steven D.		
APPLICANT: Hemmerlich, Stefan		
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3		
FILE REFERENCE: 6510-107CON		
CURRENT APPLICATION NUMBER: US/09/816,825		
PRIOR FILING DATE: 2001-03-22		
PRIOR APPLICATION NUMBER: 09/045,284		
NUMBER OF SEQ ID NOS: 9		
SOFTWARE: FastSeq for Windows Version 4.0		
SEQ ID NO 1		
LENGTH: 2032		
TYPE: DNA		
ORGANISM: Homo sapiens		
US-09-816-825-1		
Query Match	87.4%	Score 1784.8 ; DB 9 ; Length 2032 ;
Best Local Similarity	96.8%	Pred. No. 0 ;
Matches 1850 ;	Conservative	0 ; Mismatches 17 ; Indels 5 ; Gaps 5 ;
Oy	94	AAGCCCAAGCAGCAAGGTCTTCACATTCAGCAGCAATGCTACTGCTTAATAAATGAAGACT 153
Db	148	AAGCCCGCTTCAAGGCTTCACACTTCAGCAGCAATGCTACTGCTTAATAAATGAAGACT 207
Oy	154	CCTGCTGTTTCTGTTTCCAGATGCGCATCTTGCTCTATTTCTTCACATGTACAGCA 213
Db	208	CCTGCTGTTTCTGTTTCCAGATGCGCATCTTGCTCTATTTCTTCACATGTACAGCA 267
Oy	214	CAACATGCTCCCTGCTATGAAGGACAGCCGAGGAGGACGAGTGTGTTCTGTC 273
Db	268	CAACATGCTCCCTGCTATGAAGGACAGCCGAGGAGGACGAGTGTGTTCTGTC 327
Oy	274	TTCTGCGGCTCTGCGCTCTCTTTGTTGCGGACGCTTTTGGGACAGCAGATGTTT 333

```

|||||
328 TTCTGGCCCTGGCTCTCTTGTGGGGCAGCTTTTGGCCACCACCAATGTTT 397
QY 334 CTACCTGATGAGCCCGCTGGCAGCTGTGATGACCTTCAAGCAGACAGCCGCTGAT 393
Db 388 CTACCTGATGAGCCCGCTGGCAGCTGTGATGACCTTCAAGCAGACAGCCGCTGAT 447
QY 394 GCTGCACATGCTGTGCGGGATCTGATACGGGCGCTCTTCTTGTGCGACATGAGCTCT 453
Db 448 GCTGCACATGCTGTGCGGGATCTGATACGGGCGCTCTTCTTGTGCGACATGAGCTCT 507
QY 454 TGAATGCTACATGAGAACTGTGTCCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAAAC 513
Db 508 TGAATGCTACATGAGAACTGTGTCCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAAAC 567
QY 514 CCGGGCCCTGTGTGTGACACCTGTGACATCATCCACAAGATGAATATCCTCCCG 573
Db 568 CCGGGCCCTGTGTGTGACACCTGTGACATCATCCACAAGATGAATATCCTCCCG 627
QY 574 GGTCTACATGAGGCTCTGTGACAGTCAAGACGCTTTGAGTGTGAGAGAGGCTGCG 633
Db 628 GGTCTACATGAGGCTCTGTGACAGTCAAGACGCTTTGAGTGTGAGAGAGGCTGCG 687
QY 634 CTCTCTACAGCCAGTGTGCTCAAGAGAGGTGCGCTTCTTCAACCTGACGTCCCTACCC 693
Db 688 CTCTCTACAGCCAGTGTGCTCAAGAGAGGTGCGCTTCTTCAACCTGACGTCCCTACCC 747
QY 694 GCTGCTGAAAGACCCCTCCCTCAACCTGATATCGTGCACCTGTGCGGAGCCCGGCG 753
Db 748 GCTGCTGAAAGACCCCTCCCTCAACCTGATATCGTGCACCTGTGCGGAGCCCGGCG 807
QY 754 GCTGTTCCTTCCCGGAGACGACAAAGGAGATCTCATGATTTGACATGCTGAT 813
Db 808 GCTGTTCCTTCCCGGAGACGACAAAGGAGATCTCATGATTTGACATGCTGAT 867
QY 814 GGGGAGCATGAGCAAAACTCAAGAGAGAGACCAACCTACTATGTGTGAGGTGAT 873
Db 868 GGGGAGCATGAGCAAAACTCAAGAGAGAGACCAACCTACTATGTGTGAGGTGAT 927
QY 874 CTGCCAAAGCCAGCTGTGAGATCTACAAAGACATCCAGTCTTGCACAGGCTGACAGA 933
Db 928 CTGCCAAAGCCAGCTGTGAGATCTACAAAGACATCCAGTCTTGCACAGGCTGACAGA 987
QY 934 AGGCTACCTGTGTGGCGCTATGAGACCTGGCTGAGGCCCTGTGTGCCACATTCGCG 993
Db 988 AGGCTACCTGTGTGGCGCTATGAGACCTGGCTGAGGCCCTGTGTGCCACATTCGCG 1047
QY 994 AATGATGATTCGTGTGGATTTCTTGCCTCATCTTCAAGACCTGGGTGATTAACAT 1053
Db 1048 AATGATGATTCGTGTGGATTTCTTGCCTCATCTTCAAGACCTGGGTGATTAACAT 1107
QY 1054 CACCCGAGGCAAGGCGATGGGTGACACAGCTTTCACACAAATGCCAGGATGCCCTTAA 1113
Db 1108 CACCCGAGGCAAGGCGATGGGTGACACAGCTTTCACACAAATGCCAGGATGCCCTTAA 1167
QY 1114 TGTCTCCCGGCTGTGGCGCTGTGTGGCTATGAAAAAGTTTCTCGACTTCAGAAAC 1173
Db 1168 TGTCTCCCGGCTGTGGCGCTGTGTGGCTATGAAAAAGTTTCTCGACTTCAGAAAC 1227
QY 1174 CTGTGGCGATGCAATGATTTCTGTGGGCTACCGCCAGCAGATCTGAACAAGAACAGAG 1233
Db 1228 CTGTGGCGATGCAATGATTTCTGTGGGCTACCGCCAGCAGATCTGAACAAGAACAGAG 1287
QY 1234 AAACCTGTGTGATCTTCTGTACCTGAGACTGTCCCTGAGCAAAATCCACTAAGAGG 1293
Db 1288 AAACCTGTGTGATCTTCTGTACCTGAGACTGTCCCTGAGCAAAATCCACTAAGAGG 1347
QY 1294 TTGAAGAGCTTGTGCTGCACCTGTGTGACCTGATGATCTTCTGTAATGCTTGA 1353
Db 1348 TTGAAGAGCTTGTGCTGCACCTGTGTGACCTGATGATCTTCTGTAATGCTTGA 1407
QY 1354 GCCTTGCCATATCTGTGAGCTTAATCATGATGTGTGGGTATACACTGATGTGAT 1413
|||||

```

```

Db 1408 GCCTTGCCATATCTGTGAGCTTAATCATGATGTGTGGGTATACACTGATGTGAT 1467
QY 1414 TGTGTCCACAGCTGTCTCAAGCAGAGAGACTTTGTGTCCATGCTGTGTGATAGAAACAG 1473
Db 1468 TGTGTCCACAGCTGTCTCAAGCAGAGAGACTTTGTGTCCATGCTGTGTGATAGAAACAG 1527
QY 1474 ACTGGGAACTTATGTGAGCAGACACATCCACAGTGAAGAGGATTTGCTTCTTC 1533
Db 1528 ACTGGGAACTTATGTGAGCAGACACATCCACAGTGAAGAGGATTTGCTTCTTC 1587
QY 1534 TTTTCTTGATCTTCTGTGTGGGAGACTTCAAGACTTTGTGGCTGGAGGCTTATTA 1593
Db 1588 TTTTCTTGATCTTCTGTGTGGGAGACTTCAAGACTTTGTGGCTGGAGGCTTATTA 1647
QY 1594 GCACGACAGATATCAGTGAATGATTCATTAACCTCCCTGTCCATCTTCCCAATG 1653
Db 1648 GCACGACAGATATCAGTGAATGATTCATTAACCTCCCTGTCCATCTTCCCAATG 1707
QY 1654 GGGAAATGATCTTTCACAAAGAGCTCACCAGATTTTCCACAGAGATGGAATTTCTGAG 1713
Db 1708 GGGAAATGATCTTTCACAAAGAGCTCACCAGATTTTCCACAGAGATGGAATTTCTGAG 1767
QY 1714 CCCTTGAGTTCCTCAATGGGATTCAGAGAGAGAGTGGGAAACAAGTTGGATGCTACTT 1773
Db 1768 CCCTTGAGTTCCTCAATGGGATTCAGAGAGAGAGTGGGAAACAAGTTGGATGCTACTT 1826
QY 1774 ATGAGCTTGACAT-ACAGCTATCGGTAATCAGAAATAGAAACAAATCTCTG-ACAA 1831
Db 1827 ATGAGCTTGACATCAGCTATCGGTAATCAGAAATAGAAACAAATCTCTGACAA 1886
QY 1832 AGAGCAAGCTTAACTTCAAGAGTGGCTGGGCTTGAATTTCA-TTTCCCTTG 1890
Db 1887 AGAGCAAGCTTAACTTCAAGAGTGGCTGGGCTTGAATTTCACTTCCCTTG 1946
QY 1891 CATTTCCCATTAATAGAAACTTGTACCTGTGAAGTTCATCTGTTAATACTTAA 1950
Db 1947 CATTTCCCATTAATAGAAACTTGTACCTGTGAAGTTCATCTGTTAATACTTAA 2005
QY 1951 TTCCCAATTAAG 1962
Db 2006 TTCCCAATTAAG 2017

RESULT 2
US-10-007-262-2
; Sequence 2, Application US/10007262
; Publication No. US20020164748A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerlich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/10/007,262
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens
US-10-007-262-2

Query Match 87.4%; Score 1784.8; DB 13; Length 2032;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1850; Conservative 0; Mismatches 17; Indels 5; Gaps 5;

QY 94 AAGCCCAAGCAGCAAGTCTTCCACTTTCAGCACAATGTACTGCTTAAATAATGAAGCT 153
Db 148 AAGCCCGCTTTCAGAGGTCTTCCACTTTCAGCACAATGTACTGCTTAAATAATGAAGCT 207

```

OY 154 CCTGCTGTTTGGTTTCCAGATGGCATCTTGGCTATTTCTCCAGATGTACAGCA 213
 DB 208 CCTGCTGTTTGGTTTCCAGATGGCATCTTGGCTATTTCTCCAGATGTACAGCA 267
 OY 214 CAACATGAGCTCCCTGTCTATGAAGGACACGCCAGAGGATGACAGCTGTCTGTCTC 273
 DB 268 CAACATGAGCTCCCTGTCTATGAAGGACACGCCAGAGGATGACAGCTGTCTGTCTC 327
 OY 274 TTCTGCGCTGTGGCTCTTCTTTTGTGGGACACTTTTGGGACACCAATGTTTT 333
 DB 328 TTCTGCGCTGTGGCTCTTCTTTTGTGGGACACTTTTGGGACACCAATGTTTT 387
 OY 334 CTACCTGATGGAGCCCGCTGGACGTGTGATGACCTTCAAGAGAGACACCGCTGAT 393
 DB 388 CTACCTGATGGAGCCCGCTGGACGTGTGATGACCTTCAAGAGAGACACCGCTGAT 447
 OY 394 GCTGACATGAGCTGTGGGATGTGATACGGGCGCTCTTGTGTGACATGAGCTCTT 453
 DB 448 GCTGACATGAGCTGTGGGATGTGATACGGGCGCTCTTGTGTGACATGAGCTCTT 507
 OY 454 TGATGCTACATGTAACCTGTGCTCCGAGAGACATGCAAGCTCTTCAAGTGGAGAAC 513
 DB 508 TGATGCTACATGTAACCTGTGCTCCGAGAGACATGCAAGCTCTTCAAGTGGAGAAC 567
 OY 514 CCGGGCCCTGTGTCTGACACCTGCTGTGATGATCCACAGATGAATCATCCCG 573
 DB 568 CCGGGCCCTGTGTCTGACACCTGCTGTGATGATCCACAGATGAATCATCCCG 627
 OY 574 GGTCTACGACAGCTCTGTGACATCAACAGCCCTTTGAGGTGTGAGAAAGCTGTCCG 633
 DB 628 GGTCTACGACAGCTCTGTGACATCAACAGCCCTTTGAGGTGTGAGAAAGCTGTCCG 687
 OY 634 CTCTACAGCAGTGGTGTCAAGAGAGGTGCTCTTCAACCTGCACTGCTCTACCC 693
 DB 688 CTCTACAGCAGTGGTGTCAAGAGAGGTGCTCTTCAACCTGCACTGCTCTACCC 747
 OY 694 GCTGCTGAAGAACCTCTCTCAACCTGATATGCTGACCTGTGCTGGAGACCCCGGGC 753
 DB 748 GCTGCTGAAGAACCTCTCTCAACCTGATATGCTGACCTGTGCTGGAGACCCCGGGC 807
 OY 754 GCTGCTGCTGCTGCTGAGAACCCACAAAGGAGATCTCATGATGACAGCTGCTGAT 813
 DB 808 GCTGCTGCTGCTGCTGAGAACCCACAAAGGAGATCTCATGATGACAGCTGCTGAT 867
 OY 814 GGGGAGCAGTGAAGAACTCAAGAAAGAGAGACCAACCTCATGATGATGAGCAGTAT 873
 DB 868 GGGGAGCAGTGAAGAACTCAAGAAAGAGAGACCAACCTCATGATGATGAGCAGTAT 927
 OY 874 CTGCAAAAGCCAGCTGAGATCTACAAAGACCATCAGTCTTGGCCAAAGGCCCTGACAGA 933
 DB 928 CTGCAAAAGCCAGCTGAGATCTACAAAGACCATCAGTCTTGGCCAAAGGCCCTGACAGA 987
 OY 934 ACGTACCTGCTTGTGCGCTATGAGAGACCTGCTGAGACCCCTGTGGCCAGACTTCCCG 993
 DB 988 ACGTACCTGCTTGTGCGCTATGAGAGACCTGCTGAGACCCCTGTGGCCAGACTTCCCG 1047
 OY 994 AATGATGAATTCGAGGATGGAATTCCTGCCCATCTTGAAGCTGGGAGCATATACAT 1053
 DB 1048 AATGATGAATTCGAGGATGGAATTCCTGCCCATCTTGAAGCTGGGAGCATATACAT 1107
 OY 1054 CACCGAGGCAAGGAGCATGGTGAACAGCTTTCACACAATGACAGAGATGGCCCTTAA 1113
 DB 1108 CACCGAGGCAAGGAGCATGGTGAACAGCTTTCACACAATGACAGAGATGGCCCTTAA 1167
 OY 1114 TGTCTCCAGCTTGGCGCTGTCTTTCCTATGAAGAAAGTCTTCTGACTTCAGAAAGC 1173
 DB 1168 TGTCTCCAGCTTGGCGCTGTCTTTCCTATGAAGAAAGTCTTCTGACTTCAGAAAGC 1227
 OY 1174 CTGTGGGAGTGCATGATTCGTGGGCTACCGCAGCTCATGATGTAAGAAAGACAG 1233
 DB 1228 CTGTGGGAGTGCATGATTCGTGGGCTACCGCAGCTCATGATGTAAGAAAGACAG 1287

OY 1234 AACCTGTGCTGATCTTCTGTCTACCTGAGCTGTCCGAGAGAAATCCACTAAGAGG 1293
 DB 1288 AACCTGTGCTGATCTTCTGTCTACCTGAGCTGTCCGAGAGAAATCCACTAAGAGG 1347
 OY 1294 TTGAGAAAGCTTCTGCTCCACCTGTGTGACAGCTTCTGTCTGAAATGCTTCTGA 1353
 DB 1348 TTGAGAAAGCTTCTGCTCCACCTGTGTGACAGCTTCTGTCTGAAATGCTTCTGA 1407
 OY 1354 GCTTGTGCTCATGCTGAGCTTACTACATGCTGTGGGTATCACAGTGAAGT 1413
 DB 1408 GCTTGTGCTCATGCTGAGCTTACTACATGCTGTGGGTATCACAGTGAAGT 1467
 OY 1414 TGTGTCCACAGCTGCTCAAGCAGAGAGCTTGTGTGTCTGTCTGTCTGTCTGTCTGT 1473
 DB 1468 TGTGTCCACAGCTGCTCAAGCAGAGAGCTTGTGTGTCTGTCTGTCTGTCTGTCTGT 1527
 OY 1474 ACTGGGAACTTATGTGAGCAGACATCCACAGTGAAGAGGATTTGCTTCTTC 1533
 DB 1528 ACTGGGAACTTATGTGAGCAGACATCCACAGTGAAGAGGATTTGCTTCTTC 1587
 OY 1534 TTTTCTGATCTTCTGCTGTGGGAGACTTCAAGACTTTGTGGCTGGAGGCTTATTTAA 1593
 DB 1588 TTTTCTGATCTTCTGCTGTGGGAGACTTCAAGACTTTGTGGCTGGAGGCTTATTTAA 1647
 OY 1594 GCACGACACATATCAGTGAATTTGATTCATTAACCTCCCTGTCCACATCTTGCCCAATG 1653
 DB 1648 GCACGACACATATCAGTGAATTTGATTCATTAACCTCCCTGTCCACATCTTGCCCAATG 1707
 OY 1654 GGGATGATCTTTCACCAAGAGCTCACACGATTTTCCACAGATGCAATTCGTGAG 1713
 DB 1708 GGGATGATCTTTCACCAAGAGCTCACACGATTTTCCACAGATGCAATTCGTGAG 1767
 OY 1714 CCCTTGGAGTCCCAATGGGATTCAGAGAAAGAGTGGGAAACAAGTGTGATGCTCTCTT 1773
 DB 1768 CCCTTGGAGTCCCAATGGGATTCAGAGAAAGAGTGGGAAACAAGTGTGATGCTCTCTT 1826
 OY 1774 ATGAGCTTGACCAT-ACAGTATGCTGATATGAGAAATTTGAACAAATCTCTG-ACAAA 1831
 DB 1827 ATGAGCTTGACCATACAGTATGCTGATATGAGAAATTTGAACAAATCTCTGACAAA 1886
 OY 1832 AGACAGCTTCTTAAGTTACACAGAGTGGCTGTGATTTGAATATCA-TTCCCTTTG 1890
 DB 1887 AGACAGCTTCTTAAGTTACACAGAGTGGCTGTGATTTGAATATCACTTCCCTCTG 1946
 OY 1891 CATTTTCCATTAATAGAAACTTGTGACTGTGAAACTGTGCACTGTAATTAATAA 1950
 DB 1947 CATTTTCCATTAATAGAAACTTGTGACTGTGAAACTGTGCACTGTAATTAATAA 2005
 OY 1951 TTCCCAATTAAG 1962
 DB 2006 TTCCCAATTAAG 2017

RESULT 3
 US-09-919-580-194
 ; Sequence 194, Application US/09919580
 ; Patent No. US20020110832A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Pyle, Ruth
 ; APPLICANT: Xu, Jianshun
 ; APPLICANT: Secrist, Heather
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; FILE REFERENCE: 210121.552
 ; CURRENT APPLICATION NUMBER: US/09/919,580
 ; NUMBER OF SEQ ID NOS: 934
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 194
 ; LENGTH: 517
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-919-580-194

US-10-212-933-1

Query Match 5.3%; Score 108; DB 14; Length 2150;
 Best Local Similarity 51.3%; Pred. No. 1.1e-22;
 Matches 311; Conservative 0; Mismatches 280; Indels 15; Gaps 2;

QY 247 CGAGCGCATGACAGCTGCTTCTGCTTCTGCTGCGCTGCTGCTTCTTCTTCTTCTGCGCA 306
 DB 814 CAAGCGCGAGTGTGTATGTATGTCACCAAGTGGCGCTGCGCTGCTTCTTCTGCTGCA 873
 QY 307 GCTTTTGGGACAGACCGATGTTTCTACCTGATGAGAGCGCGCTGCGCTGCTGCTGAT 366
 DB 874 GCTCTTCAACAGAACCTGAGAGTGTCTTCTTCTTCTATGAGCTGTGTGCGAGTGTGCA 933
 QY 367 GACCTTCAAGACAGACCGCTGATGCTGCAATGCTGTGCGGATGTATACGCGC 426
 DB 934 AAAACGTACCCCGGGGAGCGCGTTCCCTGAGGGGAGCGCGGAGCATGTGAGCGC 993
 QY 427 CGCTTCTTGTGCGACATGAGGCTCTTGTATGCTTACATGGAACCTGTCCCGGAGACA 486
 DB 994 TCTCTACCGCTCGATCTTTCGCTTTCAGCTGATACCCCGGAGGAGTGGGGGCG 1053
 QY 487 GTCC-----AGCCTCTTCAAGTGGAGAACAGCGCGCTGTGTTCTGCAC 534
 DB 1054 CAACCTCACACCTGTGGGATCTTGGGGGAGCGCACTACAGAGTGTATGCTTCCGCGC 1113
 QY 535 TCGCTGTACATCATCCCAAGATGAATCATCCCGCGCTCACTGCAAGCTCC--T 591
 DB 1114 ACTCTGTCTGCTTACCGCAAGAGTCTCGACTGTGTGAGAGCGCGTGTGCAAAA 1173
 QY 592 GTGACGTCAACAGCCCTTGAAGGTGTGAGAGAGCGCTGCTCTTACAGCAAGTGT 651
 DB 1174 GTGCCCACTCAACCGCTTGGACGCTTCAAGAGAGGTGTGCAAGTACGACGCTGT 1233
 QY 652 GCTCAAGAGGTGCGCTTCTCAACCTGCAAGTCCCTCAACCGCTGCTCAAGACCCCTC 711
 DB 1234 TATCAAGGCGGTGGGCTTCTGATGTGCTGTGCTGTGGCGCGCTTAAAGATCCAC 1293
 QY 712 CCTCAACCTGCAATGTGCTGCTGCTGCGGAGCCCGCGCGCTTCTTCCGCGACA 771
 DB 1294 GTTGGACCTCAAGGTATCTACCTATACCTATGATCTCTGCTGTGCGACGCTCCGCA 1353
 QY 772 ACAGCAAGAGGATCTCATGATTGACAGTGCATTTGATGGGGGACAGTACGACAAA 831
 DB 1354 CCGCTGCGCTCAACGCGCTATCCGGGAAAGCTACAGGTGTGCGAAGCGGGATCCAG 1413
 QY 832 ACTCAA 837
 DB 1414 AGCCCA 1419

RESULT 11

US-09-735-705-63
 ; Sequence 63, Application US/09735705
 ; Patent No. US20020052329A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Tongtong
 ; APPLICANT: Fan, Liqun
 ; APPLICANT: Kalos, Michael D.
 ; APPLICANT: Bangur, Chaitanya S.
 ; APPLICANT: Hosken, Nancy
 ; APPLICANT: Fanger, Gary R.
 ; APPLICANT: Li, Samuel X.
 ; APPLICANT: Wang, Aljun
 ; APPLICANT: Skelky, Yasir A.W.
 ; APPLICANT: Henderson, Robert A.
 ; APPLICANT: McNeill, Patricia D.
 ; APPLICANT: Fanger, Neil
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; FILE REFERENCE: 210121.455C14
 ; CURRENT APPLICATION NUMBER: US/09/735,705
 ; CURRENT FILING DATE: 2000-12-12

NUMBER OF SEQ ID NOS: 419
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 63
 ; LENGTH: 731
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(731)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-735-705-63

Query Match 3.1%; Score 62.6; DB 9; Length 731;
 Best Local Similarity 58.2%; Pred. No. 1.3e-08;
 Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 648 TGGTGTCAAGAGGTGCGCTTCTCAACCTGACATCCCTTACCCGCTGGAAGACC 707
 DB 3 TAGTCATTAAGGTTGTGCGCTTCTGACGTGGCGCTTGGCGCACTGCTGCGAGACC 62
 QY 708 CCTCCCTCAACCTGATATGTCACCTGTGCGGAGACCCCGGCGCTGTCCGTTCC 767
 DB 63 CGGCCCTGACCTCAAGTCACTGCTGTGCTGTGATCCCGCGGAGGAGTTGAC 1122
 QY 768 GAGAACGCAAGAGGATCTCATGATGACAGTGTGTATGGGAGCATGAGC 827
 DB 123 GGATCCGCTGCGCGCACCGCTCATCTGTGAGAGGCTTACAGGTGTGGCGACGAGACC 182
 QY 828 AAAAATCA 836
 DB 183 GCGAGCTCA 191

RESULT 12

US-09-850-716A-63
 ; Sequence 63, Application US/09850716A
 ; Patent No. US20020115139A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kalos, Michael D.
 ; APPLICANT: McNeill, Patricia D.
 ; APPLICANT: Retter, Marc W.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; FILE REFERENCE: 210121.455C15
 ; CURRENT APPLICATION NUMBER: US/09/850,716A
 ; CURRENT FILING DATE: 2001-05-07
 ; NUMBER OF SEQ ID NOS: 440
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 63
 ; LENGTH: 731
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(731)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-850-716A-63

Query Match 3.1%; Score 62.6; DB 10; Length 731;
 Best Local Similarity 58.2%; Pred. No. 1.3e-08;
 Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 648 TGGTGTCAAGAGGTGCGCTTCTCAACCTGACATCCCTTACCCGCTGGAAGACC 707
 DB 3 TAGTCATTAAGGTTGTGCGCTTCTGACGTGGCGCTTGGCGCACTGCTGCGAGACC 62
 QY 708 CCTCCCTCAACCTGATATGTCACCTGTGCGGAGACCCCGGCGCTGTCCGTTCC 767
 DB 63 CGGCCCTGACCTCAAGTCACTGCTGTGCTGTGATCCCGCGGAGGAGTTGAC 1222
 QY 768 GAGAACGCAAGAGGATCTCATGATGACAGTGTGTATGGGAGCATGAGC 827
 DB 123 GGATCCGCTGCGCGCACCGCTCATCTGTGAGAGGCTTACAGGTGTGGCGACGAGACC 182

LOCATION: 719, 723, 725, 730, 731
OTHER INFORMATION: n = A,T,C or G
US-10-117-982-63

Query Match 3.1%; Score 62.6; DB 12; Length 731;
Best Local Similarity 58.2%; Pred. No. 1.3e-08;
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY	648	TCGTCCTCAAGAGAGGTGCGCTTCTTCACCTGCAGTCCCTACCCGCTGCTGAAAGACC	707
Db	3	TAGTCATAAAGGAGGTGCGCTTCTTCAGACGTGCGGCTTGGCGCCACTGCTGCGAGACC	62
QY	708	CCTCCCTCAACCTGCATATCGTGCACCTGTCCGGACCCCGGGCCGCTGCTCCCTCC	767
Db	63	CGCCCTGAGCTCAAGGTGATCACTTGTGCTGATCCCGCGGCTGGCGAGTTAC	122
QY	768	GAGAAAGCACAAGGAGAGATCTCATGATGACAGTGCATGTGATGGGGCAGCATGAGC	827
Db	123	GGATCCGCTCGGGCCACGGGCTCATCCGTGAGAGGCTTACAGGTGGTGGCAGCCGAGACC	182
QY	828	AAAAACTCA	836
Db	183	CGAGCTCA	191

Search completed: August 14, 2003, 18:00:55
Job time : 513.12 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 11:31:05 ; Search time 131.773 Seconds
(without alignments)
6843.163 Million cell updates/sec

Title: US-09-645-078-1
Perfect score: 2043
Sequence: 1 gaattccattgttggta.....tgggacccaataaaaaaa 2043

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA :
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq : *
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq : *
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq : *
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq : *
5: /cgn2_6/ptodata/1/lna/PCUS.COMB.seq : *
6: /cgn2_6/ptodata/1/lna/backfiles1.seq : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1813	88.7	2065	US-09-786-240-26	Sequence 26, Appl
2	1784.8	87.4	2032	US-09-045-284A-1	Sequence 1, Appl
3	1784.8	87.4	2032	US-09-190-911-2	Sequence 2, Appl
4	141.6	6.9	2156	US-08-899-514-1	Sequence 1, Appl
5	135.2	6.6	2190	US-09-015-188-1	Sequence 1, Appl
6	128.4	6.3	2354	US-08-655-878-1	Sequence 1, Appl
7	119.4	5.8	2409	US-09-263-023-3	Sequence 3, Appl
8	119.4	5.8	2409	US-09-471-867-3	Sequence 3, Appl
9	108	5.3	2150	US-09-263-023-1	Sequence 1, Appl
10	108	5.3	2150	US-09-471-867-1	Sequence 1, Appl
11	62.6	3.1	731	US-09-040-984-63	Sequence 63, Appl
12	62.6	3.1	731	US-09-123-912-63	Sequence 63, Appl
13	62.6	3.1	731	US-09-643-597-63	Sequence 63, Appl
14	62.6	3.1	731	US-09-480-884A-63	Sequence 63, Appl
15	62.6	3.1	731	US-09-542-615A-63	Sequence 63, Appl
16	62.6	3.1	731	US-09-606-421B-63	Sequence 63, Appl
17	42.2	2.1	570	US-09-252-991A-14204	Sequence 14204, A
18	42.2	2.1	570	US-09-252-991A-14242	Sequence 14242, A
19	40	2.0	4673	US-07-638-431-1	Sequence 1, Appl
20	40	2.0	4673	PCT-US92-00018-1	Sequence 1, Appl
21	39.6	1.9	7092	US-09-620-312D-252	Sequence 252, App
22	39.6	1.9	7269	US-09-620-312D-251	Sequence 251, App
23	39.6	1.9	20165	US-09-609-816-7	Sequence 7, Appl
24	38.4	1.9	3060	US-09-252-991A-4686	Sequence 4686, Ap
25	38.4	1.9	3455	US-09-252-991A-4757	Sequence 4757, Ap
26	38.2	1.9	2367	US-09-252-991A-11973	Sequence 11973, A
27	38.2	1.9	3264	US-09-252-991A-12013	Sequence 12013, A

28	38.2	1.9	3279	US-09-252-991A-11905	Sequence 11905, A
29	37.8	1.9	498	US-09-252-991A-5845	Sequence 5845, Ap
30	37.8	1.9	1005	US-09-252-991A-5806	Sequence 5806, Ap
31	37.4	1.8	509	US-09-594-506-9	Sequence 9, Appl
32	37.2	1.8	612	US-09-252-991A-711	Sequence 711, App
33	37.2	1.8	1488	US-09-252-991A-623	Sequence 623, App
34	37	1.8	37	US-09-045-284A-7	Sequence 7, Appl
35	37	1.8	37	US-09-190-911-7	Sequence 7, Appl
36	37	1.8	1965	US-09-178-252-26	Sequence 26, Appl
37	36.8	1.8	330	US-09-252-991A-15704	Sequence 15704, A
38	36.8	1.8	768	US-09-252-991A-15652	Sequence 15652, A
39	36.8	1.8	825	US-09-252-991A-15741	Sequence 15741, A
40	36.8	1.8	1446	US-09-252-991A-7937	Sequence 7937, Ap
41	36.8	1.8	1767	US-09-252-991A-7614	Sequence 7614, Ap
42	36.8	1.8	2724	US-09-252-991A-7543	Sequence 7543, Ap
43	36.8	1.8	5894	US-08-665-259-24	Sequence 24, Appl
44	36.8	1.8	5894	US-08-762-500-24	Sequence 24, Appl
45	36.8	1.8	6525	US-08-762-500-74	Sequence 74, Appl

ALIGNMENTS

```
RESULT 1
US-09-786-240-26
Sequence 26, Application US/09786240
Patent No. 6558935
GENERAL INFORMATION:
APPLICANT: INCYTE PHARMACEUTICALS, INC.
APPLICANT: TANG, Y. Tom
APPLICANT: CORLEY, Neil C.
APPLICANT: GUEGLER, Karl J.
APPLICANT: BAUGHN, Mariah R.
APPLICANT: LAL, Preeti
APPLICANT: YDE, Henry
APPLICANT: HILLMAN, Jennifer L.
APPLICANT: AZIMZAI, Yalda
TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
FILE REFERENCE: PR-0592 PCT
CURRENT APPLICATION NUMBER: US/09/786,240
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 09/186,779; unassigned; 60/133,6
PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1998-11-04; 1999-05-11
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PERL Program
SEQ ID NO 26
LENGTH: 2065
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
OTHER INFORMATION: incyte ID No. 6558935 2617407CBI
US-09-786-240-26
Query Match
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1902; Conservative 0; Mismatches 20; Indels 7; Gaps 7;
94 AAGCCCAAGCCACAGAGTCTTCCAGCTACGACCAATGCTACTGCTTAAAGAAAGCT 153
|||||
140 AAGCCCTGCTGACAGGCTTCCACTTCAGACCAATGCTACTGCTTAAAGAAAGCT 199
|||||
154 CCGCTGCTTCTGCTGCTTCCAGAGTGGCATTTGGCTTCTTCCAGATGTCAGCA 213
|||||
200 CCGCTGCTTCTGCTGCTTCCAGAGTGGCATTTGGCTTCTTCCAGATGTCAGCA 259
|||||
214 CAACATAGCTCCCTGCTCTATGAGGACAGCCGACGATGACGCTGCTTCTGCT 273
|||||
260 CAATATAGCTCCCTGCTCTATGAGGACAGCCGACGATGACGCTGCTTCTGCT 319
|||||
274 TTCTGCGCTGCTGCTGCTTCTTCTTGGGGGACGCTTTGGGACGACCAAGATGTTT 333
|||||
320 TTCTGCGCTGCTGCTGCTTCTTCTTGGGGGACGCTTTGGGACGACCAAGATGTTT 379
|||||
```

```

QY 334 CTACCTGATGAGACCCGCTGGACGCTGTGATGACCTTCAAGCAGACGACCGCTGAT 393
Db 380 CTACCGATGAGACCCGCTGGACGCTGTGATGACCTTCAAGCAGACGACCGCTGAT 439
QY 394 GGTGCATGAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 453
Db 440 GGTGCATGAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 499
QY 454 TATGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 513
Db 500 TATGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 559
QY 514 CCGGGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 573
Db 560 CCGGGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 618
QY 574 GGTCTCATGAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 632
Db 619 GGTCTCATGAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 678
QY 633 GCTCTCATGAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 692
Db 679 GCTCTCATGAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 738
QY 693 CCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 752
Db 739 CCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 798
QY 753 CCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 812
Db 799 CCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 858
QY 813 TGGGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 872
Db 859 TGGGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 918
QY 873 TGTGCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 932
Db 919 TGTGCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 978
QY 933 AACGCTACCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 992
Db 979 AACGCTACCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1038
QY 993 GAATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1052
Db 1039 GAATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1098
QY 1053 TCACCCGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1112
Db 1099 TCACCCGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1158
QY 1113 ATGTCTCCAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1172
Db 1159 ATGTCTCCAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1218
QY 1173 CCTGTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1232
Db 1219 CCTGTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1278
QY 1233 GAAACCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1292
Db 1279 GAAACCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1338
QY 1293 GTTGAAGAGCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1352
Db 1339 GTTGAAGAGCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1398
QY 1353 AGCCTTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1412
Db 1399 AGCCTTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1458

```

```

QY 1413 TTGTCTCACAGCTGCTCAAGCAGAAAGACTTTTGTGTCATGCTTGTCTTGA 1472
Db 1459 TTGTCTCACAGCTGCTCAAGCAGAAAGACTTTTGTGTCATGCTTGTCTTGA 1518
QY 1473 GACTGGGAAACCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1532
Db 1519 GACTGGGAAACCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1578
QY 1533 CTTTCTTATGCTTCTGCTGTGAGGAGACTTCAAGACCTTGTGCTGAGGCTTAT 1592
Db 1579 CTTTCTTATGCTTCTGCTGTGAGGAGACTTCAAGACCTTGTGCTGAGGCTTAT 1638
QY 1593 AGCAGACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1652
Db 1639 AGCAGACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1698
QY 1653 GGGGAATGATGCTTCAACAAAGAGCTCAACAGATTTTCCACAGATGCGAATCTGA 1712
Db 1699 GGGGAATGATGCTTCAACAAAGAGCTCAACAGATTTTCCACAGATGCGAATCTGA 1758
QY 1713 GGCCTTGGAGTTCCCAATGGGATTCAGAGAGAGAGTGGAAACAGTTGATGCTTACT 1772
Db 1759 GGCCTTGGAGTTCCCAATGGGATTCAGAGAGAGAGTGGAAACAGTTGATGCTTACT 1817
QY 1773 TATGACTTGACCAT-ACAAGCTATGCTTATGATGATGATGATGATGATGATGAT 1830
Db 1818 TATGACTTGACCAT-ACAAGCTATGCTTATGATGATGATGATGATGATGATGAT 1877
QY 1831 AAGAGCAAGCTCTTAAAGTACACAAAGAGTGGGCTGATTTGAATATCA-TTTTCCCTT 1889
Db 1878 AAGAGCAAGCTCTTAAAGTACACAAAGAGTGGGCTGATTTGAATATCA-TTTTCCCTT 1937
QY 1890 GCATTTTCCCATTAATAGAAACTTTGACTGTGAAACTTGCATCTGTATATCAATA 1949
Db 1938 GCATTTTCCCATTAATAGAAACTTTGACTGTGAAACTTGCATCTGTATATCAATA 1996
QY 1950 ATTCCCAATTAAGGCTCTTTAGAAATGCTCTTTTATGCTTCTTAATATAGAGTA 2009
Db 1997 ATTCCCAATTAAGGCTCTTTAGAAATGCTCTTTTATGCTTCTTAATATAGAGTA 2056
QY 2010 AATGTTTCAAT 2018
Db 2057 AATGTTTCAAT 2065

RESULT 2
US-09-045-284A-1
: Sequence 1, Application US/09045284A
: Patent No. 6265192
: GENERAL INFORMATION:
: APPLICANT: Bistrup, Annette
: APPLICANT: Rosen, Steven D.
: APPLICANT: Hemmerlich, Stefan
: TITLE OF INVENTION: GLYCOXYL SULFOTRANSFERASE-3
: FILE REFERENCE: 6510-107051
: CURRENT APPLICATION NUMBER: US/09/045, 284A
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 2032
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-045-284A-1

Query Match 87.4%; Score 1784.8; DB 3; Length 2032;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1850; Conservative 0; Mismatches 17; Indels 5; Gaps 5;
QY 94 AAGCCCAAGCAGAGGCTTCCACTTCAAGCAGCAATGCTTACTTAAATAAATGAGCT 153
Db 148 AAGCCCGCTTCAAGGCTTCCACTTCAAGCAGCAATGCTTACTTAAATAAATGAGCT 207

```


OY	154	CCGCTGTTTCTGGTTTCCCGAATGGCCATCTTGGCTCTATTCTTCCACATGTCACAGCA	213
Db	208	CTCTCTGTTTCTTGGTTTCCCGAATGGCCATCTTGGCTCTATTCTTCCACATGTCACAGCA	267
OY	214	CAACATCAGCTCCCTGTCTATGAAGGCAACGCCGAGCGCATATGCAGCGTGGATCTCTC	273
Db	268	CAACATCAGCTCCCTGTCTATGAAGGCAACGCCGAGCGCATATGCAGCGTGGATCTCTC	327
OY	274	TTCTGGCGCTCTGGCTCTTCTTTTGTGGGGCAGCTTTTGGGACGACCCAGATGTTTT	333
Db	328	TTCTGGCGCTCTGGCTCTTCTTTTGTGGGGCAGCTTTTGGGACGACCCAGATGTTTT	387
OY	334	CTACCTATGAGAGCCCGCTGGACAGTGTGATACCTTCAAGACAGACCCGCTGAT	393
Db	388	CTACCTATGAGAGCCCGCTGGACAGTGTGATACCTTCAAGACAGACCCGCTGAT	447
OY	394	GCTGCACATGCTGTGGGGATCTGATACGGGGCGTCTTCTGTGGCATGAGCGTCT	453
Db	448	GCTGCACATGCTGTGGGGATCTGATACGGGGCGTCTTCTGTGGCATGAGCGTCT	507
OY	454	TGATGCTCAATGGAACCTGTTCCCGGAGACAGTCCAGCTCTTTCAGTGGGAACAG	513
Db	508	TGATGCTCAATGGAACCTGTTCCCGGAGACAGTCCAGCTCTTTCAGTGGGAACAG	567
OY	514	CCGGGGCCCTGGTCTCTGCACCTGCTGTGATATCCCAAAATGAAATCATCCCGG	573
Db	568	CCGGGGCCCTGGTCTCTGCACCTGCTGTGATATCCCAAAATGAAATCATCCCGG	627
OY	574	GGCTCACTGCAAGGCTCCTGTGCAGTCAACACCCCTTGTAGGTGGTGGAGAAGGCTGGC	633
Db	628	GGCTCACTGCAAGGCTCCTGTGCAGTCAACACCCCTTGTAGGTGGTGGAGAAGGCTGGC	687
OY	634	CTCTCAACGCCAGTGTGTCTCAAGAGAGTGCGCTTCTTCAACCTGCAGTCCCTACCC	693
Db	688	CTCTCAACGCCAGTGTGTCTCAAGAGAGTGCGCTTCTTCAACCTGCAGTCCCTACCC	747
OY	694	GCTGCTGAAABACCCCTCCCTCAACCTGCATATGTGTGACCTGTGCGGGACCCCGGGC	753
Db	748	GCTGCTGAAABACCCCTCCCTCAACCTGCATATGTGTGACCTGTGCGGGACCCCGGGC	807
OY	754	CGTGTTCGTTCCCGAGAACGCACAAAGAGAGATCTCATGATTGACATCTCCATTTGAT	813
Db	808	CGTGTTCGTTCCCGAGAACGCACAAAGAGAGATCTCATGATTGACATCTCCATTTGAT	867
OY	814	GGGGCAGCATGAGCAAAAACTCAAGAAAGAGAGCAACCTACTATGTGTGATGAGTCTAT	873
Db	868	GGGGCAGCATGAGCAAAAACTCAAGAAAGAGAGCAACCTACTATGTGTGATGAGTCTAT	927
OY	874	CTGCCAAAGCCAGTGGAGATCTACAGAACATCTCAGTCTTGGCCAAAGGCCCTGACAGA	933
Db	928	CTGCCAAAGCCAGTGGAGATCTACAGAACATCTCAGTCTTGGCCAAAGGCCCTGACAGA	987
OY	934	ACGCTACCTGTTGTGGCTATATGAGAGACTGTGAGTGCACCCCTGTGGCCACAGACTTCCG	993
Db	988	ACGCTACCTGTTGTGGCTATATGAGAGACTGTGAGTGCACCCCTGTGGCCACAGACTTCCG	1047
OY	994	AATGTATGAATTCGTGGATTGGAATTTCTGCCCATCTTCAGACCTGTGGTCAATACAT	1053
Db	1048	AATGTATGAATTCGTGGATTGGAATTTCTGCCCATCTTCAGACCTGTGGTCAATACAT	1107
OY	1054	CACCCGAGGCAAGGGCATGGGTGACACCGCTTTCACACAATATGCCAGGATGCCCTTAA	1113
Db	1108	CACCCGAGGCAAGGGCATGGGTGACACCGCTTTCACACAATATGCCAGGATGCCCTTAA	1167
OY	1114	TGTGTCACAGGCTTGGCGCTGTGCTTTCCTCATGAAGAAAGTTCTGCATCTCGAAGAC	1173
Db	1168	TGTGTCACAGGCTTGGCGCTGTGCTTTCCTCATGAAGAAAGTTCTGCATCTCGAAGAC	1227
OY	1174	CTGTGGCATCCATGAATTTGTGGGTACCGGCACGTCAATCTGAACAAGAACAGAG	1233
Db	1228	CTGTGGCATCCATGAATTTGTGGGTACCGGCACGTCAATCTGAACAAGAACAGAG	1287
OY	1234	AAACCTGTTGCTGGATCTTCTGTCTACCTGAGCTGTCCCTGAGCAAAATCTAAGAGG	1293

```

Db      1288  AAACCTGTGGATCTTCTGTACTACGTGACCTGCTCAGCAAAATCCACTAGAGGG 1347
QY      1294  TTGAGAAAGCCTTCTGCTGCACCTGCTGTCAGCCTCAGTCACTTCTCTGATGCTTCTGA 1353
Db      1348  TTGAGAAAGCCTTCTGCTGCACCTGCTGTCAGCCTCAGTCACTTCTCTGATGCTTCTGA 1407
QY      1354  GCGTTGCTCATCATCTCTGAGCGCTTAACATACATGTCCTGGGTATCACATGAGTGTAGT 1413
Db      1408  GCGTTGCTCATCATCTCTGAGCGCTTAACATACATGTCCTGGGTATCACATGAGTGTAGT 1467
QY      1414  TGTGTCCACAGTGTCTCAAGCAAGAGACTTGTGTGCATGCTTGTGTGTCTAGAAAAAG 1473
Db      1468  TGTGTCCACAGTGTCTCAAGCAAGAGACTTGTGTGCATGCTTGTGTGTCTAGAAAAAG 1527
QY      1474  ACTGGGGAACCTTAATGTGAGCAGACACATCCACCAGTGAACACAGGATATGCTTCTTC 1533
Db      1528  ACTGGGGAACCTTAATGTGAGCAGACACATCCACCAGTGAACACAGGATATGCTTCTTC 1587
QY      1534  TTTTCTGTGATCTTCTCTGCTGTGGCAGACTTCAGAGACTTGTGTGCTGAGAGCCTTAATA 1593
Db      1588  TTTTCTGTGATCTTCTCTGCTGTGGCAGACTTCAGAGACTTGTGTGCTGAGAGCCTTAATA 1647
QY      1594  GCAGAGACACATATCAGTGAATGATGCATTAACCTGCCGTGCACATCTTGCCCATG 1653
Db      1648  GCAGAGACACATATCAGTGAATGATGCATTAACCTGCCGTGCACATCTTGCCCATG 1707
QY      1654  GGGATGATGATTTTTCACCAAAAGACTCCACAGACTTTTTCACAGAGATGCGAATTCGAG 1713
Db      1708  GGGATGATGATTTTTCACCAAAAGACTCCACAGACTTTTTCACAGAGATGCGAATTCGAG 1767
QY      1714  CCTTTGAGATTCCCAATGGATTCAAGSAGAAAGTGGGAACAAGTTGGATGCTTACTT 1773
Db      1768  CCTTTGAGATTCCCAATGGATTCAAGSAGAAAGTGGGAACAAGTTGGATGCTTACTT 1826
QY      1774  ATGAGCTTGACCAT-ACAGCATATGGGTATGAGAAATGTGAACAAAATCTCTG-ACAAA 1831
Db      1827  ATGAGCTTGACCATACAGCTATGCTGATATAGAAATGTGAACAAAATCTCTCACAAA 1886
QY      1832  AGAGCAAGCTTAAATTACACAGAGTGGCTGGCTTGAATTTGAATATCA-TTTCCTCTTTG 1890
Db      1887  AGAGCAAGCTTAAATTACACAGAGTGGCTGGCTGGCTGGCTGCAATTTGAATATCACTTCCCTCTG 1946
QY      1891  CATTTTCCCATTAACATAGAAAACCTTGAACCTGTGAACCTTCCATCTGTTAATACTAAA 1950
Db      1947  CATTTTCCCATTAACATAGAAAACCTTGAACCTGTGAAC-TCGCATCTGTTAATACTAAA 2005
QY      1951  TTCCCAATTAAG 1962
Db      2006  TTCCCAATTAAG 2017

RESULT 3
US-09-190-911-2
; Sequence 2, Application US/09190911
; Patent No. 6365365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/09/190,911
; CURRENT FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens

```

US-09-190-911-2

Query Match 87.4%; Score 1784.8; DB 4; Length 2032;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1850; Conservative 0; Mismatches 17; Indels 5; Gaps 5;

QY 94 AGCCCAACCCAGAGCTCTTCCACTTACAGACACATGCTACTGCTTAAAAAATGAAGCT 153
 DB 148 AAGCCCGCTTTGAGAGCTCTTCCACTTACAGACACATGCTACTGCTTAAAAAATGAAGCT 207
 QY 154 CCGCTGTTTCGCTTCCAGATGAGACCTTGGCTATCTTCCACATGACAGCA 213
 DB 208 CCGCTGTTTCGCTTCCAGATGAGACCTTGGCTATCTTCCACATGACAGCA 267
 QY 214 CAACATCAGCTCCCTGCTATGAGAGCAGACCCGAGCCATGACAGCTGCTGTTGTC 273
 DB 268 CAACATCAGCTCCCTGCTATGAGAGCAGACCCGAGCCATGACAGCTGCTGTTGTC 327
 QY 274 TTCCTGGCCTCTGGCTCTTCTTTTGGGGCAGCTTTTGGGAGACCCAGATGTTT 333
 DB 328 TTCCTGGCCTCTGGCTCTTCTTTTGGGGCAGCTTTTGGGAGACCCAGATGTTT 387
 QY 334 CTACCTGATGAGCCCGCTGGACGCTGAGATGACCTTCAAGCAGACAGCCGCTGAT 393
 DB 388 CTACCTGATGAGCCCGCTGGACGCTGAGATGACCTTCAAGCAGACAGCCGCTGAT 447
 QY 394 GCTGCACATGGCTGTGCGGGATCTGATACGGCCGCTTCTTGTGACATGAGCGTCT 453
 DB 448 GCTGCACATGGCTGTGCGGGATCTGATACGGCCGCTTCTTGTGACATGAGCGTCT 507
 QY 454 TGAATGCTACATGGAACCTGCTGCTCCCGAGACAGTCCAGCTTTCAGTGGAGAACAG 513
 DB 508 TGAATGCTACATGGAACCTGCTGCTCCCGAGACAGTCCAGCTTTCAGTGGAGAACAG 567
 QY 514 CCGGGCCGCTGTGCTGACCTGCTGACATCATCCCAAGATGAATCATCCCGG 573
 DB 568 CCGGGCCGCTGTGCTGACCTGCTGACATCATCCCAAGATGAATCATCCCGG 627
 QY 574 GGGCTACGACAGCTCTGTGACAGTCAACGCCCTTGAAGTGGAGAGCCCTGCG 633
 DB 628 GGGCTACGACAGCTCTGTGACAGTCAACGCCCTTGAAGTGGAGAGCCCTGCG 687
 QY 634 CTCTCTACGACAGCTGTGCTCAAGAGAGGGGCTTCAACCTGACATCCCTCAACC 693
 DB 688 CTCTCTACGACAGCTGTGCTCAAGAGAGGGGCTTCAACCTGACATCCCTCAACC 747
 Y 694 GCTGCTGAAGAAGCCCTCCCTCAACCTGATATCTGACACTGCTGCGGAGCCCGGGC 753
 DB 748 GCTGCTGAAGAAGCCCTCCCTCAACCTGATATCTGACACTGCTGCGGAGCCCGGGC 807
 QY 754 CCGTGTCCGTTCCCGAGAAAGCACAAGAGGAGATCTCATGATTTGACAGTCTGAT 813
 DB 808 CCGTGTCCGTTCCCGAGAAAGCACAAGAGGAGATCTCATGATTTGACAGTCTGAT 867
 QY 814 GGGGAGCATGAGCAAAAACTCAAGAAAGAGAGCAACCTACTATGATGAGGTCAAT 873
 DB 868 GGGGAGCATGAGCAAAAACTCAAGAAAGAGAGCAACCTACTATGATGAGGTCAAT 927
 QY 874 CTGCCAAAGCAGCTGAGATCTACAAGACATCCAGTCTTGCCTCCCAAGGCCCTGACAGA 933
 DB 928 CTGCCAAAGCAGCTGAGATCTACAAGACATCCAGTCTTGCCTCCCAAGGCCCTGACAGA 987
 QY 934 AGCTTACCTGCTTGTGCGCTATGAGAGACTGGCTGAGCCCTGTCGAGACTTCCCG 993
 DB 988 AGCTTACCTGCTTGTGCGCTATGAGAGACTGGCTGAGCCCTGTCGAGACTTCCCG 1047
 QY 994 AATGATGAATTCGTTGGGATGGAATCTTGCCTCCCTCAAGCTGCTGATGAATCAAT 1053
 DB 1048 AATGATGAATTCGTTGGGATGGAATCTTGCCTCCCTCAAGCTGCTGATGAATCAAT 1107
 QY 1054 CACCCGAGCAAGGGGATGGTGAACACGCTTTCACACAAATGCCAGGATGCCCTTAA 1113
 DB 1108 CACCCGAGCAAGGGGATGGTGAACACGCTTTCACACAAATGCCAGGATGCCCTTAA 1167

QY 1114 TGTCTCCAGAGCTTGGCGTGGTCTTGGCCATGAAAAAGTTCCTGACTTCAGAAAGC 1173
 DB 1168 TGTCTCCAGAGCTTGGCGTGGTCTTGGCCATGAAAAAGTTCCTGACTTCAGAAAGC 1227
 QY 1174 CTGTGCGATGCCATTAATTTGCTGGGCTACCGCCACGCTACATCTGAAACAGACAGAG 1233
 DB 1228 CTGTGCGATGCCATTAATTTGCTGGGCTACCGCCACGCTACATCTGAAACAGACAGAG 1287
 QY 1234 AATCCTGTTGCGATCTTCTGTCTACCTGGAATGTCCTCGAGCAATCCATCAAGAGG 1293
 DB 1288 AATCCTGTTGCGATCTTCTGTCTACCTGGAATGTCCTCGAGCAATCCATCAAGAGG 1347
 QY 1294 TTGAGAAGCTTGTGCTGCCACCTGCTGTCAGCTCAGTCACTTCTCTGAATGCTTGA 1353
 DB 1348 TTGAGAAGCTTGTGCTGCCACCTGCTGTCAGCTCAGTCACTTCTCTGAATGCTTGA 1407
 QY 1354 GCCTTCCCTACATCTGTAGACCTTAACTACATGCTGTGGGATACACACTGAGTGAAGT 1413
 DB 1408 GCCTTCCCTACATCTGTAGACCTTAACTACATGCTGTGGGATACACACTGAGTGAAGT 1467
 QY 1414 TGTGTACACAGCTGTCAAGCAGAGAGGACTTTGTGCTCATGCTGTGTGAAAAAG 1473
 DB 1468 TGTGTACACAGCTGTCAAGCAGAGAGGACTTTGTGCTCATGCTGTGTGAAAAAG 1527
 QY 1474 ACTGGGGAACCTTATGTGAGCAGACATCCACAGATGAACAGGGTATTGCTCTTCTC 1533
 DB 1528 ACTGGGGAACCTTATGTGAGCAGACATCCACAGATGAACAGGGTATTGCTCTTCTC 1587
 QY 1534 TTTTCTTGATCTTCTGCTGCTGGGACACTTTCAGAGCTTTGGCCTGAGAGCCTTAA 1593
 DB 1588 TTTTCTTGATCTTCTGCTGCTGGGACACTTTCAGAGCTTTGGCCTGAGAGCCTTAA 1647
 QY 1594 GCACGACAGATACAGTGAATGATCCATTAACCTCCCTGACATCTTGCCCAAG 1653
 DB 1648 GCACGACAGATACAGTGAATGATCCATTAACCTCCCTGACATCTTGCCCAAG 1707
 QY 1654 GGGATGATCTTTCACCAAGAGCTCACACGATTTTCCACAGAGATGCCAATCTGAG 1713
 DB 1708 GGGATGATCTTTCACCAAGAGCTCACACGATTTTCCACAGAGATGCCAATCTGAG 1767
 QY 1714 CCTTGGAGTTCCCATGGGATTCAGAGAGAGGGGAAACAAGTTGGATGCCACTT 1773
 DB 1768 CCTTGGAGTTCCCATGGGATTCAGAGAGAGGGGAAACAAGTTGGATGCCACTT 1826
 QY 1774 ATGAGCTTGACCAT-ACAGCTATCGGTAAATCAAGAAATATGAACAAATCTCTG-ACAAA 1831
 DB 1827 ATGAGCTTGACCATACAGCTATCGGTAAATCAAGAAATATGAACAAATCTCTGACAAA 1886
 QY 1832 AGAGCAAGCTTTAAGTTCAAGAGTGCCTGGGCTTGATTTGAATATCA-TTCCCTTGG 1890
 DB 1887 AGAGCAAGCTTTAAGTTCAAGAGTGCCTGGGCTTGATTTGAATATCACTTCCCTCTG 1946
 QY 1891 CATTTTCCATTAATAGAAAACCTTGACCTGGAAACTTGCAATCTGTTAATACATAAA 1950
 DB 1947 CATTTTCCATTAATAGAAAACCTTGACCTGGAAACTTGCAATCTGTTAATACATAAA 2005
 QY 1951 TTCCCAATTAAG 1962
 DB 2006 TTCCCAATTAAG 2017

RESULT 4
 US-08-899-514-1
 ; Sequence 1, Application US/08899514
 ; Patent No. 5910581
 ; GENERAL INFORMATION:
 ; APPLICANT: HABUCHI, OSAMI
 ; APPLICANT: FUKUTA, MASAKAZU
 ; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
 ; TITLE OF INVENTION: SUBOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
 ; NUMBER OF SEQUENCES: 9

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR, LLP
STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
CITY: NEWPORT BEACH
STATE: CALIFORNIA
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentlin
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899, 514
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: DANIEL E. ALTMAN
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: FOYAM21.001AUS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 714 760 0404
TELEFAX: 714 760 9502
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2156
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Human
TISSUE TYPE: Fetal brain
FEATURE:
NAME/KEY: CDS
LOCATION: 147..1583
IDENTIFICATION METHOD: S
US-08-899-514-1

Query Match      6.9%; Score 141.6; DB 2; Length 2156;
Best Local Similarity 50.5%; Pred. No. 8.2e-31;
Matches 488; Conservative 0; Mismatches 454; Indels 24; Gaps 5;

QY 257 CACGTGCTGTTGCTGCTCTCTGCGCTCTGCTCTTTTGTGGGCGACCTTTTGGG 316
DB 546 CACGTGCTGCTCATGCGCCACCGCGCTCTCTGCTGCGGAGTCTTCAAC 605
QY 317 CAGACCCAGATGTTTCTACCTGATGAGCCCGCGCTGCGACGTGTGA-----TGACC 370
DB 606 CAGACGGGACATCTTCTACCTCTGCGCGCTGCGACATCGACGACAGGTGCC 665
QY 371 TTCAAGCAGAG-----CACCGCTGATGCTGCACATGCTGTGGGAGTGTACGG 424
DB 666 TTGAGCGCGGGGGGCGCAAGCGCGGCTCGGCGCTGCTGACGGAGTGTCAAG 725
QY 425 GCCGCTTCTGTCGACATGAGCGCTTTGATGCTCAATGAACTGTGCCCGGAGA 484
DB 726 CAGCTCTTCTGTCGACCTGTAGTGTGAGACATTCACGCGCTCCCGAGGAC 785
QY 485 CAGTCCAGCCTTTCACT-----GGAGAACAGCGGGCGCTGTGTTCGACCTGCC 538
DB 786 CACTGACTGATCTGATGTCGCGGGGCTCCAGCGCTCTCTGTGAGAGACCCCTC 845
QY 539 TGTGACATCATCCACAAAGATGAATCATCCCGGGGCTCACTGAGGCTCTGTGAGT 598
DB 846 TGTACGCCCTTCTGCAAGAGGTCTTCGAGAGTACACTGCAAGAACCGCGCTGCGGC 905
QY 599 CACACGCCCTTGTGAGTGTGAGAGAGCGCTGCGCTCTCAAGCCACGAGTGTCTCAAG 658
DB 906 CCCCTCAACGTGACGTGTGCGCGCAGAGGCTGCGCGCCCAAGAGACATGAGCCCTCAAG 965

```

```

QY 659 GAGGAGCGCTTCTCAACGTCGAGCCCTTACCGCGTGTGAAGACCCCTCCCTCAAC 718
DB 966 GCGGTGCCATTCGCGACACTGTGAGTCTTCGACCGCGCTGCGGACGCCCTGAGC 1025
QY 719 CTGCATATGTCGACACTGTGTCCGGACCCCGGGCGGTTCCTGCTCCGAGAACGACA 778
DB 1026 CTGCGCGTCATTCGAGCTGTGCGGACCCCGGGCGGTGTGCTCGCATGATGTGCGC 1085
QY 779 AAGGAGATCTCATGATTGACAGTGTGATGTGGGCGACATGACAAACTCAAG 838
DB 1086 TTCCCGCGCAAGTATTAAGACCTGGAAGAGTGGCTGAGACGAGGCGGACGCGCTG 1145
QY 839 AAGGAGACCAACCTACTATGTATGTGAGGTGATCTCCAAAGGACACTGAGATCTAC 898
DB 1146 AGGGAAGAGA---GCTCAGCGGCTGCGGGGCAACTCGAGAGCATCCGCTGTCCGC 1202
QY 899 AAGACATCACTGCTTGTCCCAAGGCGCTGAGAACGCTACTGCTTGTGCGCTATGAG 958
DB 1203 GAGCTGGGGCTGCGGACCGCGCTGCTGCGGGGCGCTACATGCTGTGCGCTACGAG 1262
QY 959 GACCTGCTCGAGCCCTGTGTGCCCAAGCTTCCGAATGTATGATTTGTGGATGTGAA 1018
DB 1263 GACGTGACGCGGGGCGCTGTGAGAGGCGCGGAGATGTACCCGTTGCGCGCATCCCC 1322
QY 1019 TTCTTGCCCATCTTCAAGACCTGGGTGATTAACATCACCGAGGCAAGGCGATGGTAC 1078
DB 1323 CTGACCCCGCAGGTGTGAAGATGTGATCCAAAAGAACGACGAG---CGGCCACAGCGC 1379
QY 1079 CACGCTTTCACACAAATGCGAGGATGCCCTTAATGTCTCCAGGCTTGGCGCTGTCT 1138
DB 1380 AGCGCATCTACTCTCACGACGAAGAACTCTGAGAGCTTGAGAACTGGCGCTTCAGC 1439
QY 1139 TTGCCCCATGAAAAGGTTTCTGCACTTGTAGAAAGCTGTGGCGATGCCATTAATTTGCTG 1198
DB 1440 ATGCCCTTCAAGTGGGCCAGAGTGTGTAGGCGCCCGTGTGCGGCTGCTCATGCGCTCTTC 1499
QY 1199 GCGTAC 1204
DB 1500 GCGTAC 1505

RESULT 5
US-09-015-188-1
Sequence 1, Application US/09015188C
Patent No. 639358
GENERAL INFORMATION:
APPLICANT: Williams, Kevin J
TITLE OF INVENTION: A Human Gene Encoding Human Chondrolectin
FILE REFERENCE: JEFF-0231
CURRENT APPLICATION NUMBER: US/09/015.188C
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentlin Ver. 2.0
SEQ ID NO 1
LENGTH: 2190
TYPE: DNA
ORGANISM: Homo sapiens
US-09-015-188-1

Query Match      6.6%; Score 135.2; DB 4; Length 2190;
Best Local Similarity 49.1%; Pred. No. 5.9e-29;
Matches 507; Conservative 0; Mismatches 483; Indels 42; Gaps 4;

QY 257 CACGCGCTGTTGCTGCTCTCTGCGCTCTGCTCTTTTGTGGGCGACCTTTTGGG 316
DB 308 CACATCTCATCTCTGCGCACCGCGCAGCGGCTCTCTCTGCTGTGGGCGACCTTTCAAC 367
QY 317 CAGACCCAGATGTTTCTTCTACCTGATGAGCGCGCGCTGCGACGTGTGATGATCACTTCAAG 376
DB 368 CAGACCGTGAAGCTTCTTCTTCTGATGAGCCCTCTACGACGTTCAGAACACGCTCAATC 427

```

```

QY 377 CAGAGACGCGCTGATGCTGACATGAGCTGTGCGG-----412
DB 428 CCCCCGCTTACCCAGAGGCAAGACCCGCGGAGCGGCTCATGCTAGGCGCCAGCCG 487
QY 413 GATGTGATACGGGCGCTCTTCTTGTGCGATGAGGCTCTTGTGATCCCTACATGAACT 472
DB 488 GACCTCTCGGAGGCGCTTACGACGCTCTTACTCTCTGTGAGAACTACATCAAGCGG 547
QY 473 GGTCCCCGAGAGTCCAGCC-----TCTTCACTGGAGAAAGCGGCGCGCTGTG 526
DB 548 CCGCGGCTACACACACACACAGAGATCTTCCGCGCGGCGCGCGCGCTCTCTG 607
QY 527 TGTGACCTGCTGTGATCATCCACAGATGAATCATCC---CCGCGCTCACTGC 583
DB 608 TCCCGGCTGTGTGACACCTCCGCGGCGAGCGGAGCTGCTCTGTGAGAGGCGGAGCTG 667
QY 584 AGGCTCTGCTGATCAACAGCCCTTGTGAGTGTGAGAGAGGCGCGCTCTCTACG 643
DB 668 GTGCGGAGTGGGCTACTCAACCTGACCGCTGCGCGGAGCGGCTGCGGAGCGGAGC 727
QY 644 CAGTGTGCTCAAGAGGCTGCGCTCTTCAACCTGACCTGCTTACCGCTGTGAAA 703
DB 728 CAGTGTGCTCAAGAGGCTGCGCGCTGCGCGGAGTGAAGAGCTGCGCGCTGTGAAA 787
QY 704 GACCGCTCCCTCAAGCTGATCTGTGACCTGTGCGGAGCGCGCGCGCTGTGCGT 763
DB 788 GACCGCGGATTAACCTCAAGGCTGATCAAGCTGTGCGGAGCGCGCGCTGTGCGT 847
QY 764 TCCCGAGACGACAAAGGAGATCTCATGATGACAGTGTGATGAGGCGGAGCAT 823
DB 848 TCGCGGAGGAGACCTTCCGCGAGCTGACGCTGCGCGCTGTGATGAGGAGCGGAG 907
QY 824 GAGCAAAATCAAGAGAGAGACCAACCTTATGTATGAGAGCTCATCTGTCCAAAG 883
DB 908 AGGAAACCTTCAACCTGAGAGCTGACGAGCTGACGAGCTGTGAGAGCTCTTCAAC 967
QY 884 CAGCTGAGATCTACAGACATCCAGTCTTGTGCGCGAGCGCTGAGAGAGCTTACTG 943
DB 968 TCGGTGTCCACGCGCTCATGCGGCGCGCGCTGCTCAAGGCGAG-----TACATG 1018
QY 944 CTTGTGCTGATGAGAGCTGCTGAGCGCGCTGTGCGCGAGCTTCCGAGATGTAGAA 1003
DB 1019 TTGTGTGCTGATGAGAGCTGCTGAGCGCGCTGTGCGCGAGCTTCAAGAGAGCTTACGG 1078
QY 1004 TTGTGTGATGAGATCTTGTGCGCGCTTCAAGAGCTGCTGAGAGCTTCAAGAGAGCT 1063
DB 1079 TTCTGTGAGCTGCGCTGAGAGAGCTGCGCGCTGCTGATCCAGAAACAGCGGCGG 1138
QY 1064 AAGGGGATGGTGAACAGCTTTCACACAATGCCAGAGGATCCCTTATGTCTCCAG 1123
DB 1139 GACCCCTACCTGGGAGAGCAAAATGCGACCTGCGAAACTCGCGCGCGAGCGCGAG 1198
QY 1124 GCTTGTGCTGCTTGTGCTTGTGAAAGGTTTCTGACTTCAAGAGCTTGTGCGAT 1183
DB 1199 AAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1258
QY 1184 GCGATGAATTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1243
DB 1259 GTGCTGTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1318
QY 1244 CTGATCTTCTG 1255
DB 1319 GTGAGCTGTG 1330

```

```

RESULT 6
US-08-655-878-1
; Sequence 1, Application US/08655878
; Patent No. 5827713
; GENERAL INFORMATION:
; APPLICANT: FUKUTA, MASARAZU
; APPLICANT: HABUCHI, OSAMI

```

```

? TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
? NUMBER OF SEQUENCES: 9
? CORRESPONDENCE ADDRESS:
? ADDRESSEE:
? STREET:
? CITY:
? STATE:
? COUNTRY:
? ZIP:
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/655,878
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME:
? REGISTRATION NUMBER:
? REFERENCE/DOCKET NUMBER:
? TELECOMMUNICATION INFORMATION:
? TELEPHONE:
? TELEFAX:
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2354
? TYPE: nucleic acid
? STRANDEDNESS: both
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA to mRNA
? ORIGINAL SOURCE:
? ORGANISM: Chick
? TISSUE TYPE: Embryo chondrocyte
? FEATURE: NAME/KEY: CDS
? FEATURE: LOCATION: 211..1584
? FEATURE: IDENTIFICATION METHOD: P
? FEATURE: NAME/KEY: s1g-peptide
? FEATURE: LOCATION: 211..309
? FEATURE: IDENTIFICATION METHOD: P
? FEATURE: NAME/KEY: mat.peptide
? FEATURE: LOCATION: 310..1584
? FEATURE: IDENTIFICATION METHOD: P
? FEATURE: NAME/KEY: transmembrane domain
? FEATURE: LOCATION: 280..321
? FEATURE: IDENTIFICATION METHOD: P
? FEATURE: NAME/KEY: potential N-glycosylation site
? FEATURE: LOCATION: 394..402
? FEATURE: IDENTIFICATION METHOD: S
? FEATURE: NAME/KEY: potential N-glycosylation site
? FEATURE: LOCATION: 427..435
? FEATURE: IDENTIFICATION METHOD: S
? FEATURE: NAME/KEY: potential N-glycosylation site
? FEATURE: LOCATION: 493..501
? FEATURE: IDENTIFICATION METHOD: S
? FEATURE: NAME/KEY: potential N-glycosylation site
? FEATURE: LOCATION: 916..924
? FEATURE: IDENTIFICATION METHOD: S
? FEATURE: NAME/KEY: potential N-glycosylation site
? FEATURE: LOCATION: 1405..1413

```

```

? FEATURE: IDENTIFICATION METHOD: S
?
? FEATURE:
?
? FEATURE: NAME/KEY: potential N-glycosylation site
? LOCATION: 1537..1545
?
? FEATURE: IDENTIFICATION METHOD: S
US-08-655-878-1

```

Query Match	6.3%	Score 128.4	DB 1	Length 2354
Best Local Similarity	48.4%	Pred. No. 5.8e-27		
Matches 496	Conservative	0	Mismatches 501	Indels 27
				Gaps 4

QY	250	GGCGATGCACAGCTGTGGTCTTGTCTCTCCCTGGCGCTCTGGCTCTTCTTTTGGGGCAGCT	309
Db	543	GCGGCGGCATSTCTGCTGATGGCCACCACACGCAACCGGCTCTCTCTTCTGGGGAGTT	602
QY	310	TTTTGGGAGACACCAGATGTTTTCTTACTGATGGAGCCCGCTGGCAAGCTGTGGATGAC	369
Db	603	CTTCAACACAGGAGGCAACATATTCTACTCTTTGAGAGCCCGCTGGCAATCGAAGAC	662
QY	370	CTTCAA-----GCAGAGCACCGCCCTGGATGCTGCACATGGCTGTGGCGGATCT	417
Db	663	GGTACTCTTTGAGCCAGGGGGGGCAAGCGCGTGGGCTCGCGCCCTGGTGTACCGGAGCT	722
QY	418	GATACGGGAGCCCTTCTTGTGCGACATGAGGCTCTTATGCGCTTACATGGAACCGTGC	477
Db	723	GCTCGACAGAGTCTCTCTCTCTGCGACCTTACATTTGAGAGACTTCACTCACAGCGCC	782
QY	478	CCGGAGACAGTCCAG-----CTCTTCTCACTGGGAGAACAGCCGGGCGCTGTCTTCTGC	531
Db	783	CGAGGACACACTTACTGCTGCTCGCTTTCGGCGGGGCTCCAGCCACTCATCTGTGAGGA	842
QY	532	ACCTGCGCTGTGACATCATCCCAAGAATATCATCCCCGGGGCTCACTGAGGCTCT	591
Db	843	GCCCGCTGCAACACCCACGCTTCACAAAGAGTTTGTAGAATACCACTGCAAGAACCGCG	902
QY	592	GTCGAGTCAACAGCCCTTGTGAGTGTGGAAAGGCGCGCGCTCTTACAGCAGCTGCT	651
Db	903	CTGGGGGCTCTCAACATACAGCTGTGGCAGCTGMAAGCATCGCGGCCAAGCAGCATGGC	962
QY	652	GCTCAGAGAGGTGCGCTTCTTCAACCTGACAGTCCCTTACCGCGCTGTGAAGACCCCTC	711
Db	963	CTTMAAAGCGGTGGGCATCCGCGAGCTGGAATTTCTTCAGGCCCTCGCGGAGACCCCG	1022
QY	712	CCTCAACCTGATATPGTGCACAGCTGTCGGGAGACCCCGGGCGGTTCCTTCCCGAGA	771
Db	1023	GCTGAGCTGGGCATTTATCCAGCTGTGTGCGGAGACCACTGCGCGTGTGTCTC-----	1076
QY	772	ACGCACAAGGAGATCTCATGATATGACAGTGTGCAATTTGATGTGGGGCAGCATGAGCAAA	831
Db	1077	GGCGATGTGGCTTCTCGGGCAGTACGAGAGCTGGAAGAGTGGGGGCCAGAGGGGA	1133
QY	832	ACTCAAGAGAGGAGACCAACCTTACTATGTATGTAGTAGCTCATCTGCCAAGGCACTGGA	891
Db	1137	GGCCCCCTGTGAGGAGGACGAGGTGCACAGGCTGTGGGGGCAACTCGAGAGCATCGGCT	1196
QY	892	GATCTAAGACATCCACAGTCTTGTGCCAAGGCCCTGTGAGAACGCTACTGCTGTGG	951
Db	1197	GTCGGCGAGGTGGGACTCGCGGACCGCGCTGCTGTGGAGGCCGTTTACATGCTGTGCG	1255
QY	952	CTATGAGGACCTGCTCGAGCCCTGTGGCCCAAGCTTCCGGAATGTGATTCGTGGG	1011
Db	1257	CTAGAGGAGACTGTGCAGCGGCGCGCTGCGAAGCGCTGAGATGTACCCCTTGGCGG	1311
QY	1012	ATTGGAATTTCTGGCCCATCTTTCAGACTGTGGGTGATTAACATACCGGAGCAAGGCAAT	1071
Db	1317	CATCCACCCCAACGCCACAGGTGAGAGGTGATTCGCGCCCAACGACGAGCACCACAGA	1377
QY	1072	GGGTGACACAGCTTTCACACAAATGCCAGGAGATGCCCTTATGTCTCCACAGGCTTGGCG	1131
Db	1377	CAGCAAGGCACTTATCTACACGAGCAG--AAGAAGCTCTGGGAGCAGTTTGAAGATGGCG	1433
QY	1132	CTGGTCTTGGCCCTATGAAAAGTTTCTCGACTTCAGAAAGCTGTGGCATGCCATGAA	1191

[illegible]

RESULT 7
US-09-263-023-3
; Sequence 3, Application US/09263023

GENERAL INFORMATION:
 APPLICANT: UCHIMURA, Kenji
 APPLICANT: Muramatsu, Hideki
 APPLICANT: Kadomatsu, Kenji
 APPLICANT: Kanmagi, Reiji
 APPLICANT: Habuchi, Osami
 APPLICANT: Muramatsu, Takashi
 TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
 TYPE OF INVENTION: DNA ENCODING THE SAME
 FILE REFERENCE: TOYAMA1.001AUS
 CURRENT APPLICATION NUMBER: US/09/263,023
 CURRENT FILING DATE: 1999-03-05
 EARLIER APPLICATION NUMBER: JP 10-54007
 EARLIER FILING DATE: 1998-03-05
 EARLIER APPLICATION NUMBER: JP 10-177844
 EARLIER FILING DATE: 1998-06-24
 NUMBER OF SEQ ID NOS: 10
 SOFTWARE: FastSeq for Windows Version 3.0

```

; CDNA ID NO :
; LENGTH: 2409
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (390)...(1841)
US-09-263-023-3

```

Query Match	5.8%	Score 119.4	DB 3	Length 2409
Best Local Similarity	49.1%	Pred. No. 2.4e-24		
Matches 508; Conservative	0	Mismatches 446;	Indels 81;	Gaps 4;

OY		261	TGCGTGGTTCTGTCTTCACGCGGCCTGGCCTCTTCTTTTGAGGGACATTTTGGGCAAGC	320
Db		751	TGATCGATTACACACAGCGCGCTGTGGCTGTCTGTTCTTGGCGAGTATTCACAACGAA	810
OY		321	ACCAGATGTTTTCTACTGATGGAAGCCGCCCTGGCACGTGTGATGTACTTCACAGAGA	380
Db		811	ATCCGAGAGTCTTCTTCTCTACGAGCCAGTAGTGGCATGTATGGCAAAAACGTATCCGG	870
OY		381	GCACGCCCTGGATCTCGACATGGCTGTGGGGATCTGATACGGGCGCTCTTGTGCGG	440
Db		871	GGGAGCGCGTTTCCCTGCAGGGGGGACGCGGGGACATCTAAGCGCTCTTAACGCGTAGC	930
OY		441	ACAGAGCGCTTTTGATGCTCCACATGGAACCTGTGCCCCGGGAGAC-----AGT	488
Db		931	ACCTCTCTCTTCCATGTTATATAGCCCCGGGGGAGGGGGGCGAACCTCACACAGC	990
OY		489	CCAGCCTCTTTCAGATGGGAGAAACGCGGCGCTGTCTTCTGACACCTGCTGTGACATCA	548
Db		991	TGGGCATCTTGGGGGACGACCAACCAAGATGTGTGCTGTGCACACACTGTGCCCGGCT	1050
OY		549	TCCCACAAGATGAATACTATCCCGGGGCTCACTGAGGCTC---CTGTGCAGTCAACAGC	605
Db		1051	ACCGCAAGAGAGGTGTGGGGATTGGTGGACAGCCGTTGCCAAGAAAGTCCCGCCACAGC	1110
OY		606	CCTTGAAGTGTGGAGAAAGGCGTCCGCGCTCTTACAGCCAGCTGTGCTCAAGAGAGTGC	665
Db		1111	GCCGTGGGCGCTTTCGAGAGAGGAGTGGCCGCAAGTACACCGCACTATGTCTATTAAGGGGTGGC	1170

QY 535 TGCCTGTGATCATTCACCAAGATGAATCATCCCGGGCTCACTGCAGGCTC- --T 591
Db 1114 ACTGTCTGCTGCTACCGCAAGAGGTCTCGGACTGTGAGACGACCGGTGTGCAAAA 1173
QY 592 GTGACGTCAACAGCCCTTTGAGGTGTGTGAGAGAGCCCTCCCTCTACAGCCACGTGT 651
Db 1174 GTGCCACCTCAACGCTGTGACGCTTCGAGAGAGGTGTGCAAGTACCGCACGCTGTGT 1233
QY 652 GCTCAAGAGAGTGGCTTCTTCAACCTGCAGTCCCTCAACCGCTCTGTAAGACCCCTC 711
Db 1234 TATCAAGGGCGTGGGGGTCTTGATGTGGCTGTGTGTGGGGCGCTCTTAAGATGTACG 1293
QY 712 CCTCAACCTGCATATGTGTGCTGCTCCGGAGACCCGGGCGCTTCCGTCCGAGAG 771
Db 1294 CTTGACCTCAAGGTATCATCTACCTAGTATCTGCTGTGCTGTGCCAGCTCCGCGCAT 1353
QY 772 AGCAGCAAGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 831
Db 1354 CCGCTCGCTCAACGCGCTCATCCGGGAAAGCTTACAGGTGTGTGCGAAGCCGGGATCAAG 1413
QY 832 ACTCAA 837
Db 1414 AGCCCA 1419

RESULT 11

US-09-040-984-63
; Sequence 63, Application US/09040984
; Patent No. 6210883
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS
; TITLE OF INVENTION: OF LUNG CANCER
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,984
; FILING DATE: 18-MAR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.456
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-282-6031
; TELEX:
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 731 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-040-984-63

Query Match 3.1%; Score 62.6; DB 3; Length 731;
Best Local Similarity 58.2%; Pred. No. 3.7e-08;
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
648 TGGTCTCAAGAGGTGGCGCTTCTTCAACCTGCAGTCCCTCTACCGCTGTGTAAGAGACC 707

Db 3 TACTCATTAAGGGGTGTGCGGCTTCTGACGTGGCGGTCTTGGGCCCACTGCTCGAGACC 62
QY 708 CTTCCCTCAACGCTCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 767
Db 63 CGGCGCTGACCTCAAGGTATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122
QY 768 GAGACGCAACAAAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 827
Db 123 GATCCGCTGCGCGCCACAGGCTCATCTGAGAGCTTACAGGTGTGTGCGAGCGAGACC 182
QY 828 AAAACTCA 836
Db 183 GCGAGCTCA 191
RESULT 12
US-09-123-912-63
; Sequence 63, Application US/09123912A
; Patent No. 6312695
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
; FILE REFERENCE: 210121.455C1
; CURRENT APPLICATION NUMBER: US/09/123,912A
; CURRENT FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: 09/040,802
; PRIOR FILING DATE: 1998-03-18
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 63
; LENGTH: 731
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (236)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (249)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (263)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (288)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (312)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (317)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (323)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (326)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (337)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (352)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (362)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (370)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base

[illegible]

1	OTHER INFORMATION: Where n is a, c, g or t
2	NAME/KEY: modified_base
3	LOCATION: (615)
4	OTHER INFORMATION: Where n is a, c, g or t
5	NAME/KEY: modified_base
6	LOCATION: (627)
7	OTHER INFORMATION: Where n is a, c, g or t
8	NAME/KEY: modified_base
9	LOCATION: (632)
10	OTHER INFORMATION: Where n is a, c, g or t
11	NAME/KEY: modified_base
12	LOCATION: (640)
13	OTHER INFORMATION: Where n is a, c, g or t
14	NAME/KEY: modified_base
15	LOCATION: (641)
16	OTHER INFORMATION: Where n is a, c, g or t
17	NAME/KEY: modified_base
18	LOCATION: (644)
19	OTHER INFORMATION: Where n is a, c, g or t
20	NAME/KEY: modified_base
21	LOCATION: (654)
22	OTHER INFORMATION: Where n is a, c, g or t
23	NAME/KEY: modified_base
24	LOCATION: (660)
25	OTHER INFORMATION: Where n is a, c, g or t
26	NAME/KEY: modified_base
27	LOCATION: (665)
28	OTHER INFORMATION: Where n is a, c, g or t
29	NAME/KEY: modified_base
30	LOCATION: (671)
31	OTHER INFORMATION: Where n is a, c, g or t
32	NAME/KEY: modified_base
33	LOCATION: (678)
34	OTHER INFORMATION: Where n is a, c, g or t
35	NAME/KEY: modified_base
36	LOCATION: (692)
37	OTHER INFORMATION: Where n is a, c, g or t
38	NAME/KEY: modified_base
39	LOCATION: (697)
40	OTHER INFORMATION: Where n is a, c, g or t
41	NAME/KEY: modified_base
42	LOCATION: (698)
43	OTHER INFORMATION: Where n is a, c, g or t
44	NAME/KEY: modified_base
45	LOCATION: (699)
46	OTHER INFORMATION: Where n is a, c, g or t
47	NAME/KEY: modified_base
48	LOCATION: (704)
49	OTHER INFORMATION: Where n is a, c, g or t
50	NAME/KEY: modified_base
51	LOCATION: (705)
52	OTHER INFORMATION: Where n is a, c, g or t
53	NAME/KEY: modified_base
54	LOCATION: (712)
55	OTHER INFORMATION: Where n is a, c, g or t
56	NAME/KEY: modified_base
57	LOCATION: (714)
58	OTHER INFORMATION: Where n is a, c, g or t
59	NAME/KEY: modified_base
60	LOCATION: (717)
61	OTHER INFORMATION: Where n is a, c, g or t
62	NAME/KEY: modified_base
63	LOCATION: (718)
64	OTHER INFORMATION: Where n is a, c, g or t
65	NAME/KEY: modified_base
66	LOCATION: (719)
67	OTHER INFORMATION: Where n is a, c, g or t
68	NAME/KEY: modified_base
69	LOCATION: (723)

Query Match 3.1%; Score 62.6; DB 4; Length 731;
Best Local Similarity 58.2%; Pred. No. 3.7e-08;
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 648 TGGTGTCTCAAGGAGGCGCTTCTTCAACCTGCAGTCCCTACCCGCTGCTGAAGACC 707
DB 3 TAGTCATAAAGGTGTGCGCCGCTTCTGACGTGGCGCTGCGCCAGCTGCGAGACC 62
QY 708 CCTCCCTCAACCTGCATATCTGACACCTGTCGGGACCCCGGGCGCTGTTCCGTTCCC 767
DB 63 CGGCCCTGACCTCAAGGTCAATCCTGTTGGTATCCCGCGCGGTGGCGAGTTTAC 122
QY 768 GAGAAGCGCAAAAGGAGATCTCATGATGATGACAGTCCGATTGTGATGGGCGAGCATGAGC 827
DB 123 GGATCCGCTCGCGCCACGGCTCATCTCGAGAGCCTACAGGTGTGTCGCGCAGCCGAGACC 182
QY 828 AAAAAGCTCA 836
DB 183 GCGAGCTCA 191

RESULT 13
US-09-643-597-63
Sequence 63, Application US/09643597
Patent No. 6426072
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Fan, Liqun
APPLICANT: Kalos, Michael D.
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Hosken, Nancy
APPLICANT: Fanger, Gary R.
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aljun
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Henderson, Robert A.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.455C11
CURRENT APPLICATION NUMBER: US/09/643,597
CURRENT FILING DATE: 2000-08-21
NUMBER OF SEQ ID NOS: 369
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 63
LENGTH: 731
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(731)
OTHER INFORMATION: n = A,T,C or G
US-09-643-597-63

Query Match 3.1%; Score 62.6; DB 4; Length 731;
Best Local Similarity 58.2%; Pred. No. 3.7e-08;
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 648 TGGTGTCTCAAGGAGGCGCTTCTTCAACCTGCAGTCCCTACCCGCTGCTGAAGACC 707
DB 3 TAGTCATAAAGGTGTGCGCCGCTTCTGACGTGGCGCTGCGCCAGCTGCGAGACC 62
QY 708 CCTCCCTCAACCTGCATATCTGACACCTGTCGGGACCCCGGGCGCTGTTCCGTTCCC 767
DB 63 CGGCCCTGACCTCAAGGTCAATCCTGTTGGTATCCCGCGCGGTGGCGAGTTTAC 122
QY 768 GAGAAGCGCAAAAGGAGATCTCATGATGATGACAGTCCGATTGTGATGGGCGAGCATGAGC 827
DB 123 GGATCCGCTCGCGCCACGGCTCATCTCGAGAGCCTACAGGTGTGTCGCGCAGCCGAGACC 182
QY 828 AAAAAGCTCA 836
DB 183 GCGAGCTCA 191

RESULT 14
US-09-480-884A-63
Sequence 63, Application US/09480884A
Patent No. 6482597
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Fan, Liqun
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Hosken, Nancy A.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
FILE REFERENCE: 210121.455C6
CURRENT APPLICATION NUMBER: US/09/480,884A
CURRENT FILING DATE: 2001-08-27
NUMBER OF SEQ ID NOS: 330
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 63
LENGTH: 731
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(731)
OTHER INFORMATION: n = A,T,C or G
US-09-480-884A-63

Query Match 3.1%; Score 62.6; DB 4; Length 731;
Best Local Similarity 58.2%; Pred. No. 3.7e-08;
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 648 TGGTGTCTCAAGGAGGCGCTTCTTCAACCTGCAGTCCCTACCCGCTGCTGAAGACC 707
DB 3 TAGTCATAAAGGTGTGCGCCGCTTCTGACGTGGCGCTGCGCCAGCTGCGAGACC 62
QY 708 CCTCCCTCAACCTGCATATCTGACACCTGTCGGGACCCCGGGCGCTGTTCCGTTCCC 767
DB 63 CGGCCCTGACCTCAAGGTCAATCCTGTTGGTATCCCGCGCGGTGGCGAGTTTAC 122
QY 768 GAGAAGCGCAAAAGGAGATCTCATGATGATGACAGTCCGATTGTGATGGGCGAGCATGAGC 827
DB 123 GGATCCGCTCGCGCCACGGCTCATCTCGAGAGCCTACAGGTGTGTCGCGCAGCCGAGACC 182
QY 828 AAAAAGCTCA 836
DB 183 GCGAGCTCA 191

RESULT 15
US-09-542-615A-63
Sequence 63, Application US/09542615A
Patent No. 6518256
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Fan, Liqun
APPLICANT: Kalos, Michael D.
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Hosken, Nancy A.
APPLICANT: Fanger, Gary R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
FILE REFERENCE: 210121.455C8
CURRENT APPLICATION NUMBER: US/09/542,615A
CURRENT FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 350
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 63
LENGTH: 731
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:

```

; NAME/KEY: misc_feature
; LOCATION: (1)...(731)
; OTHER INFORMATION: n = A,T,C or G
US-09-542-615A-63

```

```
Query Match      3.1%; Score 62.6; DB 4; Length 731;
Best Local Similarity 58.2%; Pred. No. 3.7e-08;
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0
```

QY	648	TGGTGCCTAAGAGAGTGGCTTCTTCAACCTCAGACCCCTTAACCGGTGAGAAAGAC	707
Db	3	TAGTCATTAAGGGTGTGCGCGCTTTCGACGTGGCGGTCTTGGCGCCACTGCTGCGAGACC	62
QY	708	CCTCCCTCAACCTGCATATCGTCACCTGGTCCGGGACCCCGGGCCGCTTCCGTTCCC	767
Db	63	CGGCGCTTGAGCTCAAGGTGTCATCCACTTGGTCTCGTGATCCCGCGGGGTGGGGAGTTCAAC	122
QY	768	GAGAACGCACAAAGGAGATCTCATGATTTGACAGTCGCAATTGTGATGGGGCAGCATGAGC	827
Db	123	GGATTCGCGTCCGCGCCACGGCCCTCATCGGTGAGAGGCTTACAGGTGGTGGCCAGCCGAGACC	182
QY	828	AAAAACTCA	836
Db	183	GGGAGCTCA	191

Search completed: August 14, 2003, 17:44:28
Job time : 137.773 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 11:17:45 ; Search time 3864.66 Seconds
(without alignments)
12848.224 Million cell updates/sec

Title: US-09-645-078-1
Perfect score: 2043
Sequence: 1 gaattccatgctgttggtga.....tgggaccccaaaaaaaaaa 2043

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_estl:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vtl:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rpd:*
26: em_gss_png:*
27: em_gss_vrl:*
28: gb_gssl:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	815.6	39.9	954	12	BI823850 603039012
2	760.6	37.2	1923	11	AK009113 Mus muscu
3	647.6	31.7	668	9	AL709927 DRFZp860
4	581.4	28.5	583	2	BSM082755 Bx490456 Homo sapi

Result No.	Score	Query Match	Length	ID	Description
5	577	28.2	669	13	BU684395 UI-CF-ENO
6	507	24.8	593	12	BM969292 UI-CF-ENO
7	499.8	24.5	571	12	BM129080 UI-CF-ENO
8	496.2	24.3	680	10	BF878439 MKO-ET010
9	476	23.3	553	12	BM128370 UI-CF-ENO
10	476	23.3	553	12	BM128370 UI-CF-ENO
11	385.2	18.9	783	12	BM128831 UI-CF-ENO
12	368.6	18.0	695	12	BM969298 UI-CF-ENO
13	364.4	17.8	852	12	BM969298 UI-CF-ENO
14	364.4	17.8	852	12	BM969298 UI-CF-ENO
15	363.8	17.8	852	12	BM969298 UI-CF-ENO
16	357.8	16.5	419	9	AM572390 UI-CF-ENO
17	325.6	15.9	902	13	BM947021 UI-CF-ENO
18	288	14.1	362	9	AM002418 UI-CF-ENO
19	282.6	13.8	834	13	BM084877 UI-CF-ENO
20	274.4	13.4	571	12	BM245312 UI-CF-ENO
21	271.2	13.3	695	10	BE857538 UI-CF-ENO
22	263.6	12.9	735	10	BE857538 UI-CF-ENO
23	253.8	12.4	692	14	BY711808 UI-CF-ENO
24	252.4	12.4	620	9	AT824100 UI-CF-ENO
25	239.2	11.7	507	12	BM246681 UI-CF-ENO
26	236	11.6	804	13	BU135348 UI-CF-ENO
27	235	11.5	1067	29	CNS03K87 UI-CF-ENO
28	221	10.8	525	10	BF197521 UI-CF-ENO
29	217.8	10.7	904	13	BU241318 UI-CF-ENO
30	216.8	10.6	773	13	BU458165 UI-CF-ENO
31	213.8	10.5	768	13	BU342120 UI-CF-ENO
32	209.2	10.2	703	14	CD218830 UI-CF-ENO
33	209	10.2	803	14	BY708952 UI-CF-ENO
34	200.2	9.8	417	9	AM572510 UI-CF-ENO
35	192.6	9.4	849	29	CNS04QFN UI-CF-ENO
36	192.4	9.4	692	13	BU256241 UI-CF-ENO
37	192.2	9.4	687	14	CB167190 UI-CF-ENO
38	167	8.2	568	9	AI115260 UI-CF-ENO
39	158.8	7.8	657	29	AG035205 UI-CF-ENO
40	156.6	7.7	555	4	BX514916 UI-CF-ENO
41	156.2	7.6	517	10	BF042384 UI-CF-ENO
42	156.2	7.6	517	10	BF042384 UI-CF-ENO
43	155.2	7.6	433	28	AZ405100 UI-CF-ENO
44	154.4	7.6	543	13	BU382276 UI-CF-ENO
45	147.2	7.2	380	12	BM433243 UI-CF-ENO

ALIGNMENTS

RESULT 1
BI823850
LOCUS
DEFINITION
603039012F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5179826 5',
mRNA sequence.
ACCESSION
BI823850
VERSION
BI823850.1 GI:15935400
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
EST.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
NIH-MGC http://mgc.ncl.nih.gov/
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapds@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L14M1448 row: h column: 03
High quality sequence stop: 856.

FEATURES
source

Location/Qualifiers
1. .954
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5179826"
/lab_host="DH10B"
/clone_11b="NH_MGC_115"
/note="Organ: pooled brain, lung, testis; Vector: PCMV-SPORT6; Site: 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH_MGC Library."

BASE COUNT 199 a 297 c 250 g 208 t

ORIGIN

Query Match 39.9%; Score 815.6; DB 12; Length 954;
Best Local Similarity 97.8%; Pred. No. 8.1e-205;
Matches 933; Conservative 0; Mismatches 9; Indels 12; Gaps 10;

51 GCGCATGGCCGGCTGACAGTACAGCCCTCAAAAGCAGGAGGAGCCCAAGCCACAGG 110
|||||
1 GCGCATGGCCGGCTGACAGTACAGCCCTCAAAAGCAGGAGGAGGAGCCCAAGCCACAGG 60

111 TCTTCACCTTTCAGCAATAGTACTGCTTAAATGAGTCCCTGCTTCTGCTTT 170
|||||
61 TCTTCACCTTTCAGCAATAGTACTGCTTAAATGAGTCCCTGCTTCTGCTTT 120

171 CCCAGATGGCCATCTTGGCTATCTTTCACATGATACGCCACAACATCAGCTCCCTGT 230
|||||
121 CCCAGATGGCCATCTTGGCTATCTTTCACATGATACGCCACAACATCAGCTCCCTGT 180

231 CATTAAGGACACAGCCGAGCCGATGACAGTGTGTTGTCTTCTGCGCTGCTGCT 290
|||||
181 CATTAAGGACACAGCCGAGCCGATGACAGTGTGTTGTCTTCTGCGCTGCTGCT 240

291 CTTCTTTTGTGGGAGCTTTTGGGACAGCCAGATGTTTTCACATGATGAGAGCCG 350
|||||
241 CTTCTTTTGTGGGAGCTTTTGGGACAGCCAGATGTTTTCACATGATGAGAGCCG 300

351 CCTGCACGTGTGATGATCCTTCAAGCAGAGCAGCCGCTGATGCTGACATGGCTGTGC 410
|||||
301 CCTGCACGTGTGATGATCCTTCAAGCAGAGCAGCCGCTGATGCTGACATGGCTGTGC 360

411 GGGATCTGATAGGGCCGCTCTCTGTGTGACATGAGCCTTTGATGCTTACATGGAAC 470
|||||
361 GGGATCTGATAGGGCCGCTCTCTGTGTGACATGAGCCTTTGATGCTTACATGGAAC 420

471 CTGTGTCGCCGAGACAGTCCAGCTCTTTCAGTGGAGAAACAGCGGCGCTGTGTTCG 530
|||||
421 CTGTGTCGCCGAGACAGTCCAGCTCTTTCAGTGGAGAAACAGCGGCGCTGTGTTCG 480

531 CACCTGCTGTGATCATCCACAAGATGAATATATCCCGGGGCTCAGTCAGAGCTCC 590
|||||
481 CACCTGCTGTGATCATCCACAAGATGAATATATCCCGGGGCTCAGTCAGAGCTCC 540

591 TGTGAGTCAACAGCCCTTGAAGTGTGGAGAGGCGCTGCGCTCTCTACACCAAGCTGG 650
|||||
541 TGTGAGTCAACAGCCCTTGAAGTGTGGAGAGGCGCTGCGCTCTCTACACCAAGCTGG 600

651 TGTCTAAGAGAGTGGCTTCTTCAACATGACATCCCTTACCCGCTGTGAAGAGCCCT 710
|||||
601 TGTCTAAGAGAGTGGCTTCTTCAACATGACATCCCTTACCCGCTGTGAAGAGCCCT 660

711 CCTCTAAGCTGATATGCTGACCC-TGCTCCGGAGACCCCGGGCGGTGTTCGTTCCGCA 769
|||||
661 CCTCTAAGCTGATATGCTGACCTTGTGTCGGAGACCCCGGGCGGTGTTCGTTCCGCA 720

QY 770 GAACGCA-CAAGGAGATCTCATGATTGA-CAGTCGATTTGATGGGCGACATGAGC 827
|||||

DB 721 GAAGCAGCAACCAAGGAGATCTCATGATTGATACCGATGCTGATGGGCGACATGAGC 780
|||||

QY 828 AAAAA--CTCAGAAAGGAGGAGACC-AACCTACTATGTATG-CAGTCACTGCCAAA-- 881
|||||

DB 781 AAAAACTCAGAAAGGAGGAGACCAAACTTACTATGTATGTCAGAGTCTATGCCAAGAAG 840
|||||

QY 882 GCGAGCTGAGATCTTAC-NAGACATCCAGTCTTGCCCAAGGCGCTGAGGAGAGCTAC 940
|||||

DB 841 CCAGGCTGAGATCTTCAAAAGACATCCAGTCTTGCCCAAGGCGCTGAGGAGAGCTTAC 900
|||||

QY 941 CTGCTTGTGCG-CTATGAGGAGCTGCTGCA-GCCCTGTGCGCCAGACTTCCC 992
|||||

DB 901 CTGCTTGTGCGCAATGAGAGACTGCTGCTGAGGAGGCGCTGTGCGCCAGACTTCCC 954
|||||

RESULT 2
AK009113
LOCUS
DEFINITION
MUS musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310003G18 product:carbohydrate (chondroitin 6-sulfate) sulfotransferase 4, full insert sequence.
AK009113
VERSION
AK009113.1 GI:12843701
KEYWORDS
HTC; CAP trapper.
SOURCE
MUS musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
MEDLINE
PUBMED
10349636

TITLE
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
2049374
MEDLINE
PUBMED
11042159

REFERENCE
AUTHORS
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichannel sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
MEDLINE
PUBMED
11078661

TITLE
4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aikawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Iizawa, N., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiya, H., Kuehl, P., Lewis, S., Matsuo, Y., Nkaido, I., Pesole, G., Quackenbush, J., Schriml, L.M., Stabli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Bonfield, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustinich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombert, P., Noridone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,


```

Db      981 GATTTTGGCCACCTCCAAACATGGGTTTACATGTACCCGCGGCAAGGCGATGGT 1040
OY      1076 GACCACGCTTTCACACAATGCGAGGATGCGCTTATGTCTCCAGGCTGGGCTGG 1135
Db      1041 CAGCATGCTTTCATCTAAGCCGAGAACGCCCTCAAGCTCTCTAGGCGTGGCTTGG 1100
OY      1136 TCTTGGCCTATGAAAAGGTTTCTGACATTCAGAAAACCTGTGGCGATCCATGAATTG 1195
Db      1101 TCTTACCTTACGAAAAGGTTTCCGAGCTTCAAGATCGCTGGGCTAGCTATGATTTG 1160
OY      1196 CTGGGCTACCGCAGCTAGATCTGAAACAAGAACGAACTGTGTGATCTTCTG 1255
Db      1161 CTGGGATACCTCCAGCTCAGATCTCAACAAGAACGAACTGTGTGATCTTCTG 1220
OY      1256 TCTACCTGAGCTGTCCCTGAGCAATCACTAAGAGGTTGAGAAAGCTTGTGCTG-CCA 1313
Db      1221 TCTCTCTCCCATATCTTGGGGAGGCTTCCGAGAAAGTTAAGAGGCTGTCTGACCC 1280
OY      1314 CCTGGGTGAGCTCAGTCACTTCTCTAATGCTTCTGAGCCTTGCATCATCTGTAG 1373
Db      1281 CTTGGTCCAGCCTTAGTAC--CATTAACGCGACAGAAAGCTTAAAGTATTAACCAAC 1337
OY      1374 CTTAATCATATGTCTGTGGTATCAGCTAGTGTGATGTGTGTCACAGCTGTCAAG 1433
Db      1338 TGAAGTCCCTTCTTCTCCTCAGCCGAGAGAGGCTTGTGTCTATCTCATGTCTA 1397
OY      1434 CAGAAAGACTTGTGTGTCAGCTGTGTCTAGAAAACAGACTGGGAACTTATGTAG 1493
Db      1398 CCTTACACTGAGCCTTAAGAAAGAAAGATCTTGTCTTGAATACTTAGG 1457
OY      1494 CAGCAGATCCCAACGATGAACAGGTA--TTGCTTCTTCTTCTTCTTGTATCTTCTG 1550
Db      1458 AACCTTAAGACGCTTGTGACCTGTCAAGCAAGACTTCTGTATCTTGTGCTTCA 1517
OY      1551 TCTGGGACACTTCAAGACTTGTGGCCTGGAGCCTTAAAGCAGCAGCATATCTAG 1610
Db      1518 CCGTGCATACCTTGGAGACTAGGCTGTGGAGGATACGAGACAGCAACAGCATCTG 1577
OY      1611 TGAATTTATCAATAAAGCTCCCTGTCCATCTTGGCCATGGGGAATGATCTTTCAC 1670
Db      1578 TGAAGTGTCTGTAAACCTCCCTGTCCATCTTCCGAAAGAAAGGGAATAAATA 1637
OY      1671 CAAAGAGCTCAGCAGATTTTCCACAGATGGA-----ATTGTGAGCCCT 1717
Db      1638 GCCAAGAAAGCGGCTTTCACCAAGAGCTCCCAAACTCCACAGAGCATTAACCC 1697
OY      1718 TGAAGTTCCTCAATGGGATTCAGAAAGAAAGTGGAAAGTGTGGATCTTCTATGA 1777
Db      1698 TCAAGTTTCCAGAGGATTTGAGAAAGAGGAGCGGGGTGGGTGCTTATCT-CTGAATC 1756
OY      1778 GCTTACCATACAGCTATGCTATGCAAAATATGAAACAAATCTCTGACAAAGAGCA 1837
Db      1757 TCTGCCATTTGACATATGATATCAGAAACAGAGCAAA--CAACACAAACAGAGGC 1813
OY      1838 AGCTCTTAGTTACACAAGGTGGCTGGCTTGAAT 1870
Db      1814 AAGCTCAAGTTCAGAGCGTGTGTGATGAT 1846

```

```

RESULT 3
AL709927 668 bp mRNA 1linear EST 12-JUN-2003
LOCUS DKEP68602364_r1.686 (synonym: hlcc3) Homo sapiens cDNA clone
DEFINITION DKEP68602364 5', mRNA sequence.
ACCESSION AL709927
VERSION AL709927.1 GI:19693282
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 668)
AUTHORS Bahr, A., Lauber, J., Mewes, H. W., Well, B., Amlid, C., Osanger, A., Fodo

```

```

TITLE      'G., Han, M. and Wiemann, S.
JOURNAL    EST (Bahr, A., Lauber, J., Mewes, H. W., Well, B., et al.)
UNPUBLISHED
COMMENT    Contact: Bahr A
MIPS
Ingolstaedter Landstr. 1, D-85764 Neuberg, Germany
This is the 5' sequence of the clone insert
from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by Otago (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.
No ST sequence available.
This clone (DKEP68602364) is available at the RZPD in Berlin.
Please contact the RZPD: Reesourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
1. 668
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKEP68602364"
/tissue_type="human skeletal muscle"
/ab_host="DH10B"
/clone_lib="R2PD (synonym: hlcc3)"
/notes="Vector: pTriplex2; Site 1: SfiI; Site 2: SfiI;
cDNA-collection"
BASE COUNT 132 a 202 c 178 g 154 t 2 others
ORIGIN
Query Match 31.7%; Score 647.6; DB 9; Length 668;
Best Local Similarity 99.1%; Pred. No. 2,6e-160;
Matches 661; Conservative 0; Mismatches 5; Indels 1; Gaps 1;
OY      36 GGGTAGAGAGAAAAGCGATGGCCGCGTAGAGTAGAGCTCTCAAAAGCAGCAGGA 95
Db      2 GGGAGAGAGAGAAAAGCGATGGCCGCGTAGAGTAGAGCTCTCAAAAGCAGCAGGA 61
OY      96 GCCAAGCAGCAAGGCTTCCACTTACAGCAATGCTACTGCTTAAATAATGAAGCTCC 155
Db      62 GCCAAGCAGCAAGGCTTCCACTTACAGCAATGCTACTGCTTAAATAATGAAGCTCC 121
OY      156 TGTCTTGTGCTTCCAGATGGCCATCTGTGCTTATTTCTCCATGTACAGCCACA 215
Db      122 TGTCTTGTGCTTCCAGATGGCCATCTGTGCTTATTTCTCCATGTACAGCCACA 181
OY      216 ACATGAGCTCCGCTATGAAAGGACAGCCGAGGAGCAAGCAGTGTCTCTCTT 275
Db      182 ACATGAGCTCCGCTATGAAAGGACAGCCGAGGAGCAAGCAGTGTCTCTCTT 241
OY      276 CTTGGGCTGTGGCTCTTCTTGTGGGAGCTTTTGGGAGCAGCCAGATGTTTCT 335
Db      242 CTTGGGCTGTGGCTCTTCTTGTGGGAGCTTTTGGGAGCAGCCAGATGTTTCT 301
OY      336 ACCTGATGAGAGCCGCTGGCAGCTGTGATGATCTTCAAGAGAGACACCGCTGATGC 395
Db      302 ACCTGATGAGAGCCGCTGGCAGCTGTGATGATCTTCAAGAGAGACACCGCTGATGC 361
OY      396 TGCATAGGCTGTGGGATCTGATACAGGAGGCTTCTTGTGCGAGATGAGGCTTTTG 455
Db      362 TGCATAGGCTGTGGGATCTGATACAGGAGGCTTCTTGTGCGAGATGAGGCTTTTG 421
OY      456 ATGCTTACATGGAACCTGTGCCGAGAGCAGTCCAGCTTTTCAAGTGGGAGACAGCC 515
Db      422 ATGCTTACATGGAACCTGTGCCGAGAGCAGTCCAGCTTTTCAAGTGGGAGACAGCC 481
OY      516 GGGCCCTGTGTCTGAGCTGCTGTGACATATATCCCAAGATGAATATCTCCCGG-6 574
Db      482 GGGCCCTGTGTCTGAGCTGCTGTGACATATATCCCAAGATGAATATCTCCCGG 541
OY      575 GCTCAGTGAAGGCTCTGAGTCAAGAGCCCTTGTGAGTGTGAGAGAGGCTGCGC 634
Db      542 GCTCAGTGAAGGCTCTGAGTCAAGAGCCCTTGTGAGTGTGAGAGAGGCTGCGC 601

```



```

QY 635 TCCTACGACGCTGCTGCTCAAGAGAGTGGCTTTCTTCAACCTGCAGTCCCTTACCG 694
DB 602 TCTACAGCCACGCTGCTGCTCAAGAGAGTGGCTTTCTTCAACCTGCAGTCCCTTACCG 661
QY 695 CTGCTGA 701
DB 662 CTGCTGA 668

RESULT 4
HSM082755 standard; RNA; EST; 583 BP.
ID HSM082755 standard; RNA; EST; 583 BP.
AC BX490456;
SV BX490456.1
XX 09-MAY-2003 (Rel. 75, Created)
DT 09-MAY-2003 (Rel. 75, Last updated, Version 1)
XX Homo sapiens mRNA; EST DKFZp686G0680_r1 (from clone DKFZp686G0680)
XX EST; expressed sequence tag.
XX Homo sapiens (human)
XX Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
XX Eutheria; Primates; Catarrhini; Homnidae; Homo.
XX [1]
XX Ansoyge W., Krieger S., Regiert T., Rittmuller C., Schwager B.,
XX Mewes H.W., Well B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.:
XX Submitted (07-MAY-2003) to the EMBL/GenBank/DBJ databases.
XX LIPS, Ingolstaedter Landstr.1, D-85764 Neuberg, GERMANY
XX
XX This is the 5' sequence of the clone insert
XX Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
XX Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
XX Heidelberg/Germany) within the cDNA sequencing consortium of
XX the German genome project.
XX No s1 sequence available.
XX This clone (DKFZp686G0680) is available at the RZPD in Berlin.
XX Please contact the RZPD: Ressourcezentrum, Heubnerweg 6,
XX 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX
XX Key Location/Qualifiers
XX source 1. 583
XX /db_xref="taxon:9606"
XX /mol_type="mRNA"
XX /organism="Homo sapiens"
XX /clone="DKFZp686G0680"
XX /DH10B; sites SfiI + SfiIb"
XX /dev_stage="adult"
XX /tissue_type="CDNA-collection"
XX
XX Sequence 583 BP; 120 A; 172 C; 154 G; 137 T; 0 other;
Query Match 28.5%; Score 581.4; DB 2; Length 583;
Best Local Similarity 99.8%; Pred. No. 9, 2e-143;
Matches 582; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 41 GAGGAGAAAGCGCATGGCCGCTAGCACTGAGCCTCTCAAAAGCAGAGGAAGCCA 100
DB 1 GAGGAGAAAGCGCATGGCCGCTAGCACTGAGCCTCTCAAAAGCAGAGGAAGCCA 60
QY 101 AGCCACAGGCTCTTCCCTCAGCAGCAATGCTACTGCTTAAATAATGAAGCTCTCTG 160
DB 61 AGCCACAGGCTCTTCCCTCAGCAGCAATGCTACTGCTTAAATAATGAAGCTCTCTG 120

```

```

QY 161 TTTCGTTTCCGATGGCCATCTTGCTCTATTCTTCCACATGTACAGCACATC 220
DB 121 TTTCGTTTCCGATGGCCATCTTGCTCTATTCTTCCACATGTACAGCACATC 180
QY 221 AGCTCCCTGTATGAAGGACACAGCCGAGCGCATGACAGTGTCTTCTCTCTG 280
DB 181 AGCTCCCTGTATGAAGGACACAGCCGAGCGCATGACAGTGTCTCTCTCTG 240
QY 281 CGCTGTGCTCTCTTCTTGTGGGAGCCTTTTGGGAGCACCAGATTTTCTACCTG 340
DB 241 CGCTGTGCTCTCTTCTTGTGGGAGCCTTTTGGGAGCACCAGATTTTCTACCTG 300
QY 341 ATGAGCGCCGCTGACAGCTGTGATACCTTCAAGAGAGCAGCCGCTGATCTCTAC 400
DB 301 ATGAGCGCCGCTGACAGCTGTGATACCTTCAAGAGAGCAGCCGCTGATCTCTAC 360
QY 401 ATGAGCGCGGATGATACAGGCGCTCTTGTGCGACATGACGCTTGTATGCC 460
DB 361 ATGAGCGCGGATGATACAGGCGCTCTTGTGCGACATGACGCTTGTATGCC 420
QY 461 TACATGACCTGCTCCCGGAGACAGTCCAGCCTTTTCACTGGAGAGACAGCCGGCC 520
DB 421 TACATGACCTGCTCCCGGAGACAGTCCAGCCTTTTCACTGGAGAGACAGCCGGCC 480
QY 521 CTGCTGTGACCTGCTGACATGATCCACAGTGAATCATCCCGGGCTCAC 580
DB 481 CTGCTGTGACCTGCTGACATGATCCACAGTGAATCATCCCGGGCTCAC 540
QY 581 TGCAGGCTCTGTGACATGACAGCCCTTGTAGGTGTGAGAG 623
DB 541 TGCAGGCTCTGTGACATGACAGCCCTTGTAGGTGTGAGAG 583

RESULT 5
B0684395 669 bp mRNA linear EST 07-Oct-2002
LOCUS B0684395/c
DEFINITION UI-CF-ENO-aco-f-08-0-UI.s1 UI-CF-ENO Homo sapiens cDNA clone
ACCESSION B0684395
VERSION B0684395.1 GI:23537302
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 669)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLES Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDIUM 97044477
PUBMED 8889548
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
CDNA library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Genetics (www.resgen.com) or from Open Biosystems
Seq primer: M13 FORWARD
POLY-A-Yes.
FEATURES
source location/Qualifiers
1. 669
/organism="Homo sapiens"
/mol_type="mRNA"

```

/db_xref="taxon:9606"
 /clone="UI-CF-ENO-aco-f-08-0-UI"
 /tissue_type="Primary Lung Cystic Fibrosis Epithelial Cells"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-CF-ENO"
 /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-CF-ENO is a cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGGT.
 TAG_TISSUE=Human Lung Epithelial Cell lines untreated IPS
 6hr to IPS 24h
 TAG_SEQ=CTGCTCAGGT"

BASE COUNT 191 a 138 c 147 g 193 t
 ORIGIN

Query Match 28.2%; Score 577; DB 13; Length 669;
 Best Local Similarity 97.6%; Pred. No. 1.4e-141;
 Matches 649; Conservative 0; Mismatches 10; Indels 6; Gaps 6;

QY 1382 ACATGCTGTGGGTATCATCAGTGTGATGTTGTGCACAGTGTCTCAAGCAGAGA 1441
 DB 669 ACATGCTGTGGGTATCATCAGTGTGATGTTGTGCACAGTGTCTCAAGCAGAGA 611
 QY 1442 CTTTGTGTGCATGCTGTGTGTAGAAACAGACCTGGGAACCTTATGAGAGACACAT 1501
 DB 610 CTTTGTGTGCATGCTGTGTGTAGAAACAGACCTGGGAACCTTATGAGAGACACAT 551
 QY 1502 CCCACAGTGAACAGAGGATATGCTCTCTCTTCTTGTGATCTCTCTCTGGGAGAC 1561
 DB 550 CCCACAGTGAACAGAGGATATGCTCTCTCTTCTTGTGATCTCTCTCTGGGAGAC 491
 QY 1562 TTCAGAGCTTGTGGCTGTGAGGCTTATTAAGCAGACAGTATCACTGGAATGATC 1621
 DB 490 TTCAGAGCTTGTGGCTGTGAGGCTTATTAAGCAGACAGTATCACTGGAATGATC 431
 QY 1622 CATTAACCTCCCTGTCACATCTTGCCCAATGGGAATGATCTTCCACAAAGACTCA 1681
 DB 430 CATTAACCTCCCTGTCACATCTTGCCCAATGGGAATGATCTTCCACAAAGACTCA 371
 QY 1682 CCAGCATTTTCCACAGAGATGCGAATCTGAGCCCTTGGAGTCCCATGAGTCAAGG 1741
 DB 370 CCAGCATTTTCCACAGAGATGCGAATCTGAGCCCTTGGAGTCCCATGAGTCAAGG 312
 QY 1742 AAGGAAGTGGGAACAAGTGTGATGCTTATGAGTGAACCAT-ACAGTATCGTGA 1800
 DB 311 AAGGAAGTGGGAACAAGTGTGATGCTTATGAGTGAACCATGAGTATCGTGA 252
 QY 1801 ATCAGAAATATGAACAAATCTCTG-ACAAAAAGCAAGCTCTTAAGTTCACAGGTGC 1859
 DB 251 ATCAGAAATATGAACAAATCTCTG-ACAAAAAGCAAGCTCTTAAGTTCACAGGTGC 192
 QY 1860 CTGGGCTGATTTGAATATCA-TTTCCTTTCATTTTCCATACATAGAAACTTTGA 1918
 DB 191 CTGGGCTGATTTGAATATCACTTCCCTCTGATTTTCCATACATAGAAACTTTGA 132
 QY 1919 CCTGTGAACCTTGCATCTGTTAATATAATATCCCAATAAGTCTGTTAGAAATCT 1978
 DB 131 CCTGTGAACCTTGCATCTGTTAATATAATATCCCAATAAGTCTGTTAGAAATCT 73
 QY 1979 CCCTTTTATGCTTCTTAATATTAAGCAGTAAATGTTCTTTTATGGATCTTAATAA 2038

DB 72 CCCTTTTATGCTTCTTAATATTAAGCAGTAAATGTTCTTTATGGATCTTAATAA 13
 QY 2039 AAAA 2043
 DB 12 AAAA 8

RESULT 6
 BM969292/c
 LOCUS
 DEFINITION
 UI-CF-ENO-accp-1-21-0-UI s1 UI-CF-ENO Homo sapiens cDNA clone
 UI-CF-ENO-accp-1-21-0-UI s1, mRNA sequence.
 BM969292
 BM969292.1 GI:19586879
 EST.
 SOURCE
 Homo sapiens (human)
 ORGANISM

REFERENCE
 AUTHORS
 TITLE
 Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL
 MEDLINE
 PUBMED
 COMMENT
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 8889548
 Contact: McCray, PB
 McCray Lab
 University of Iowa
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
 Tel: 319 356 4866
 Fax: 319 356 7171
 Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com) or from Open Biosystems (www.openbiosystems.com).
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES
 source

Location/Qualifiers
 1..593
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-CF-ENO-accp-1-21-0-UI"
 /tissue_type="Primary Lung Cystic Fibrosis Epithelial Cells"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-CF-ENO"
 /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-CF-ENO is a cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGGT.
 TAG_LIB=UI-CF-ENO
 TAG_TISSUE=Human Lung Epithelial Cell lines untreated IPS
 6hr to IPS 24h
 TAG_SEQ=CTGCTCAGGT"

BASE COUNT 157 a 127 c 137 g 172 t

ORIGIN

Query Match 24.8% Score 507; DB 12; Length 593;
 Best Local Similarity 97.4%; Pred. No. 4.9e-123;
 Matches 568; Conservative 0; Mismatches 10; Indels 5; Gaps 5;

QY 1382 ACATGCTGTGGATGATCAGACAGTGTGAGTTGTGTCACACAGCTGCTCAAGAGAGA 1441
 DB 593 ACATGCTGTGGATGATCAGACAGTGTGAGTTGTGTCACACAGCTGCTCAAGAGAGA 534
 QY 1442 CTTTGTGTCCAGCTGTGTGTCTAGAAAACAGACTGGGGAACCTTATGTAGCAGCAT 1501
 DB 533 CTTTGTGTCCAGCTGTGTGTCTAGAAAACAGACTGGGGAACCTTATGTAGCAGCAT 474
 QY 1502 CCCACAGTGAACAGGATATGCTCTTCTTCTTGTATGCTTCCGCTGGGAGAC 1561
 DB 473 CCCACAGTGAACAGGATATGCTCTTCTTCTTGTATGCTTCCGCTGGGAGAC 414
 QY 1562 TTCAGACATTTGTGGCTGTGGAGCCCTTAAAGCAGACAGATATCAGTGAATGATC 1621
 DB 413 TTCAGACATTTGTGGCTGTGGAGCCCTTAAAGCAGACAGATATCAGTGAATGATC 354
 QY 1622 CATTAACCTCCCTGTCCACATCTTGCCCAATGGGGAATGATCTTCAACCAAGAGCTCA 1661
 DB 353 CATTAACCTCCCTGTCCACATCTTGCCCAATGGGGAATGATCTTCAACCAAGAGCTCA 294
 QY 1682 CCAGCATTTTCCAGAGATGCGAATCTGTGAGCCCTTGGAGTCCCAATGGGATTCAGG 1741
 DB 293 CCAGCATTTTCCAGAGATGCGAATCTGTGAGCCCTTGGAGTCCCAATGGGATTCAGG 235
 QY 1742 AAGGAAGTGGGAAACAGGTTGGATGCTTACTTATGAGCTTGACCAT-ACAGCTATCGGTA 1800
 DB 234 AAGGAAGTGGGAAACAGGTTGGATGCTTACTTATGAGCTTGACCATGAGCTATCGGTA 175
 QY 1801 ATCGAATATGAAACAAATCTCTG-ACAAAAGCAGAGCTTAACTTACAGAGTGC 1859
 DB 174 ATCGAATATGAAACAAATCTCTG-ACAAAAGCAGAGCTTAACTTACAGAGTGC 115
 QY 1860 CTGGGCTGTGATTTGAATATCA-TTCCCTTGTGATTTTCCCATATGAAATCTTGA 1918
 DB 114 CTGGGCTGTGATTTGAATATCA-TTCCCTTGTGATTTTCCCATATGAAATCTTGA 55
 QY 1919 CCGTGAAGCTGGCATCTGTAACTTAAATCCCAATTA 1961
 DB 54 CCGTGAAGCTGGCATCTGTAACTTAAATCCCAATTA 13

RESULT 7
 BM129080 571 bp mRNA linear EST 12-MAR-2002

DEFINITION 117c04.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
 CNA clone IMAGE:567655 5' similar to TR:09Y5R3 09Y5R3
 N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE. ; mRNA sequence.

ACCESSION BM129080
 VERSION BM129080.1 GI:17123632
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 571)
 AUTHORS Melton,D., Brown,T., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
 Lemishka,I., Searles,M., Brestelli,J., Gradwohl,G., Clifton,S.,
 Hillier,L., Marra,M., Page,D., Wylie,T., Martin,J., Blisstein,A.,
 Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
 'M., Gibbons,M., McCann,R., Cole,R., Tsagarelis,V., Williams,T.,
 Jackson,Y., and Bowers,Y.
 Endocrine Pancreas Consortium

TITLE Unpublished
 JOURNAL Other ESTs: 117c04.x1
 COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute

FEATURES

source
 1..571
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:567655"
 /sex="Both"
 /tissue_type="Islets of Langerhans"
 /dev_stage="Adult"
 /lab_host="DH10B"
 /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
 /note="Organ: Pancreas; Vector: pSPORT1; Site: 1; Site 2: Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT

109 a 169 c 143 g 150 t

ORIGIN
 Query Match 24.5% Score 499.8; DB 12; Length 571;
 Best Local Similarity 96.6%; Pred. No. 4e-121;
 Matches 504; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 94 AAGCCAAAGCCAGAGGCTCTCCACTTCAGCAATGCTACTGCTTAAATAAGAGCT 153
 DB 61 AAGCCCGCTTGCAAGGCTTCCACTTCAGCAATGCTACTGCTTAAATAAGAGCT 120
 QY 154 CCGTCTGTTTCTGCTTCCAGATGGCCATCTTGCTTATCTTCCATGTAACAGCA 213
 DB 121 CCGTCTGTTTCTGCTTCCAGATGGCCATCTTGCTTATCTTCCATGTAACAGCA 180
 QY 214 CAATATAGCTCCCTGTCTATGAGAGGACACCCGAGGAGATGACAGCTGCTGCTC 273
 DB 181 CAATATAGCTCCCTGTCTATGAGAGGACACCCGAGGAGATGACAGCTGCTGCTC 240
 QY 274 TTCTGGGAGCTGCTGCTCTTCTTGTGGGAGAGTGTGGAGACCCAGATGTTTT 333
 DB 241 TTCTGGGAGCTGCTGCTCTTCTTGTGGGAGAGTGTGGAGACCCAGATGTTTT 300
 QY 334 CTACCTGATGAGCCCGGCTTGACAGTGTGATGACCTTCAAGAGAGACCGCTGAT 393
 DB 301 CTACCTGATGAGCCCGGCTTGACAGTGTGATGACCTTCAAGAGAGACCGCTGAT 360
 QY 394 GCTGCAATGCTGCTGCGGATGATGATGAGGCGGCTCTTCTTGTGGCAGATGAGCTTT 453
 DB 361 GCTGCAATGCTGCTGCGGATGATGATGAGGCGGCTCTTCTTGTGGCAGATGAGCTTT 420
 QY 454 TGATGCTTACATGGAACCTGCTCCCGAGAGAGTCCAGCTCTTTCAGTGGAGAACAG 513
 DB 421 TGATGCTTACATGGAACCTGCTCCCGAGAGAGTCCAGCTCTTTCAGTGGAGAACAG 480
 QY 514 CCGGCGCTGTGCTGACCTGCTGATCATCCCAAGATGAATCATCCCGC 573

Db 481 CGGGGCCCTGTTCTGCACTGCTGATCATCCACAAAGATGAAATCATCCCGC 540

QY 574 GGCTCACTGACAGCTCTGTCAGTCAACAG 604
 |||||
 Db 541 GGCTCACTGACAGCTCTGTCAGTCAACAG 571

RESULT 8
 BF878439 680 bp mRNA linear EST 17-JAN-2001
 LOCUS MR0-ET0109-191100-002-h06 ET0109 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF878439
 VERSION BF878439.1 GI:12268569
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 680)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Britones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H., Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL MEDLINE
 PUBMED 10737800
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704932
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0&cl=MR0-ET0109-191100-002-h06&ts=2000-11-19&td=1)
 Seg primer: puc 18 forward
 High quality sequence start: 17
 High quality sequence stop: 641.
 Location/Qualifiers
 1. 680
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="ET0109"
 /note="Organ: Lung, tumor; Vector: puc18; Site: 1: SmaI; Site: 2: SmaI; A mini-library was made by cloning products derived from ORESNES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 168 a 164 c 163 g 184 t 1 others

ORIGIN

Query Match 24.3%; Score 496.2; DB 10; Length 680;
 Best Local Similarity 93.5%; Pred. No. 3.7e-120;
 Matches 572; Conservative 0; Mismatches 34; Indels 6; Gaps 5;

QY 1328 CAGTCACTTCTGATGCTTCTGAGCCTTGCT-ACATCCTGAGCCTTAATCACTG 1386
 |||||
 Db 17 CAGTCACTTCTGATGCTTCTGAGCCTTGCTGAGCCTTAATCACTG 76
 |||||

QY 1387 TCTGTGGTATCACT--GAGTGTGATGTGTCACAGAGTCTCAAGAGAGACTT 1444
 |||||
 Db 77 TCTGTGGTATCACTGAGTGTGATGTGTCACAGTCTCAAGAGAGACTT 136
 |||||

QY 1445 TTGTGTCATGCTTGTGTGTAGAAAAACAGACTGGGGAACCTTATGTGACACACATCC 1504
 |||||
 Db 137 TTGTGTCATGCTTGTGTGTAGAAAAACAGACTGGGGAACCTTATGTGACACACATCC 196
 |||||

QY 1505 ACCAGTGAACAGGGTATTGCTTCTTCTTTTGTGATCTTCTGCGGACACTTC 1564
 |||||
 Db 197 ACCAGTGAACAGGGTATTGCTTCTTCTTTTGTGATCTTCTGCGGACACTTC 256
 |||||

QY 1565 AGAGACTTGTGGCCGTGGAGGCTTTAAGCAGACAGATGATGATGATTCAT 1624
 |||||
 Db 257 AGAGACTTGTGGCCGTGGAGGCTTTAAGCAGACAGATGATGATGATTCAT 316
 |||||

QY 1625 AAACCTCCCTGTCACATCTTCCCAATGGGAATGGATCTTTCACAAAGAGCTACCA 1684
 |||||
 Db 317 AAACCTCCCTGTCACATCTTCCCAATGGGAATGGATCTTTCACAAAGAGCTACCA 376
 |||||

QY 1685 GCATTTTCCACAGAGATGCGAATTTGAGCCCTTGAGTTCCTCCATGGATTCAGGAG 1744
 |||||
 Db 377 GCATTTTCCACAGAGATGCGAATTTGAGCCCTTGAGTTCCTCCATGGATTCAGGAG 435
 |||||

QY 1745 GAAGTGGGAACAAGTTGGATGCTTACTATGAGCTTGACATACAGTATGCTATCA 1804
 |||||
 Db 436 GAAGTGGGAACAAGTTGGATGCTTACTATGAGCTTGACATACAGTATGCTATCA 495
 |||||

QY 1805 GAATATGAAACAAATCTCTG-ACAAAAGACAGCTTAAAGTTCACAGAGTGCCTGG 1863
 |||||
 Db 496 GAATATGAAACAAATCTCTGACAAAAGACAGCTTAAAGTTCACAGAGGTCGCTGG 555
 |||||

QY 1864 GCTTATTTGAATATCA-TTCCCTTTTGATTTTCCATTAATAGAAAATTTGACCTG 1922
 |||||
 Db 556 GGTGCAATTTGATATCACTTCCCTGATTTTCCATCATAGAAAGACTTGCACCTG 615
 |||||

QY 1923 TGAACCTTGCA 1934
 |||||
 Db 616 TGAAGCTGCA 627
 |||||

RESULT 9
 BM128370 553 bp mRNA linear EST 12-MAR-2002
 LOCUS lfi3d12.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
 DEFINITION cDNA clone IMAGE:5676335 3', mRNA sequence.
 ACCESSION BM128370
 VERSION BM128370.1 GI:17122922
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 553)
 Melton, D., Brown, J., Kently, G., Permutt, A., Lee, C., Kaestner, R., Lemishka, I., Seacae, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Merra, M., Page, D., Wylie, T., Martin, J., Blustein, A., Schmitt, A., Theising, B., Ritzer, E., Honko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarelshvili, R., Williams, T., Jackson, Y. and Bowers, Y.
 Endocrine Pancreas Consortium
 Unpublished
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohpc.harvard.edu
 Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Juliana Brown (brownjefas.harvard.edu) This sequence now available from the IMAGE consortium, for clone orders contact: info@image.llnl.gov
 High quality sequence start: 425.

FEATURES
sourceLocation/Qualifiers
1. .553

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:567635"

/sex="Both"

/tissue_type="Islets of Langerhans"

/dev_stage="Adult"

/lab_host="DH10B"

/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"

/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1; Site_2: Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation: average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 163 a 111 c 127 g 152 t
ORIGINQuery Match 23.3%; Score 476; DB 12; Length 553;
Best Local Similarity 97.3%; Pred. No. 8.1e-115;

Matches 537; Conservative 0; Mismatches 10; Indels 5; Gaps 5;

QY 1486 TATGTGAGCAGACATCCACAGTGAAGAGGATATGCTCTCTTTCTTGATCT 1545

Db 553 TATGTGAGCAGACATCCACAGTGAAGAGGATATGCTCTCTTTCTTGATCT 494

QY 1546 TCTGTCTGGCAGACTTCAGAGACTTTGGCCTGAGGCGCTATTAGACGACACAGT 1605

Db 493 TCTGTCTGGCAGACTTCAGAGACTTTGGCCTGAGGCGCTATTAGACGACACAGT 434

QY 1606 ATCAGTGAATGATCCATAAAGCTCCCTGTCACATCTTCCCAATGGGAATGATCT 1665

Db 433 ATCAGTGAATGATCCATAAAGCTCCCTGTCACATCTTCCCAATGGGAATGATCT 374

QY 1666 TTCACAAAGAGCTCACAGCATTTTCCACAGAGATGCAATTCGAGCCCTGGAGTTC 1725

Db 373 TTCACAAAGAGCTCACAGCATTTTCCACAGAGATGCAATTCGAGCCCTGGAGTTC 314

QY 1726 CCAATGGGATTCAGGAAGAGTGGGAACAAGTTGGATCCTACTTATGAGCTTGACC 1785

Db 313 CCAATGGGATTCAGGAAGAGTGGGAACAAGTTGGATCCTACTTATGAGCTTGACC 255

QY 1786 AT-ACAGCTATCGTATCAGAAATATGAAACAAATCTCTG-ACAAAAGACAGCTCT 1843

Db 254 ATCAGCTATCGTATCAGAAATATGAAACAAATCTCTG-ACAAAAGACAGCTCT 195

QY 1844 TAACTTCAAGAGTGGCTGGGCTGATTTGAATATCA-TTTCCTTTGCAATTTCCCAAT 1902

Db 194 TAACTTCAAGAGTGGCTGGGCTGATTTGAATATCA-TTTCCTTTGCAATTTCCCAAT 135

QY 1903 ACATAGAAACTTGTACCTGTGAAGTGCATCTGTATTAATTAATTTCCCAATAG 1962

Db 134 ACATAGAAACTTGTACCTGTGAAGTGCATCTGTATTAATTAATTTCCCAATAG 76

QY 1963 GTTCTGTTAGAAATGCTCTTTTATGCTTCTTAATTAATTAATTAATTTCCCAAT 2022

Db 75 GTTCTGTTAGAAATGCTCTTTTATGCTTCTTAATTAATTAATTAATTTCCCAAT 16

QY 2023 ATGGGATCTTAA 2034

Db 15 ATGGGATCTTAA 4

RESULT 10
BM128831/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Endocrine Pancreas Consortium

Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,

MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@hobp.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:

Washington University Genome Sequencing Center for information on

obtaining a clone please contact: Juliana Brown

(brownj@fas.harvard.edu) This sequence now available from the IMAGE

consortium, for clone orders contact: info@image.llnl.gov

High quality sequence stop: 425.

FEATURES
sourceLocation/Qualifiers
1. .553

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5676655"

/sex="Both"

/tissue_type="Islets of Langerhans"

/dev_stage="Adult"

/lab_host="DH10B"

/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"

/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1; Site_2: Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation: average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 163 a 111 c 127 g 152 t
ORIGINQuery Match 23.3%; Score 476; DB 12; Length 553;
Best Local Similarity 97.3%; Pred. No. 8.1e-115;

Matches 537; Conservative 0; Mismatches 10; Indels 5; Gaps 5;

QY 1486 TATGTGAGCAGACATCCACAGTGAAGAGGATATGCTCTCTTTCTTGATCT 1545

Db 553 TATGTGAGCAGACATCCACAGTGAAGAGGATATGCTCTCTTTCTTGATCT 494

QY 1546 TCTGTCTGGCAGACTTCAGAGACTTTGGCCTGAGGCGCTATTAGACGACACAGT 1605

Db 1546 TCTGTCTGGCAGACTTCAGAGACTTTGGCCTGAGGCGCTATTAGACGACACAGT 1605

```

Db      493 TCCTGCTGGGACACTTACAGACATCTTGTGGCTGGAGGCGCTATTAAAGCAGACAGT 434
QY      1606 ATCAGTGAATGATCCATTAACCTCCCTGTCACATCTTGCCCAATGGGAGATGATCT 1665
Db      433 ATCAGTGAATGATCCATTAACCTCCCTGTCACATCTTGCCCAATGGGAGATGATCT 374
QY      1666 TTCACCAAGAGCTCACCGCATTTTCCACAGATGCGAATTTGAGCCCTTGGAGTTC 1725
Db      373 TTCACCAAGAGCTCACCGCATTTTCCACAGATGCGAATTTGAGCCCTTGGAGTTC 314
QY      1726 CCAATGGATTCAGAGAGAGTGGAGCAAGTGGATGGATGCTTACTTATGAGCTTAC 1785
Db      313 CCAATGGATTCAGAGAGAGTGGAGCAAGTGGATGGATGCTTACTTATGAGCTTAC 255
QY      1786 AT-ACAGCTATGCTTATCAGAAATATGAACAAATCTCTG-ACAAAGAGCAGCTCT 1843
Db      254 ATCAGCTATGCTTATCAGAAATATGAACAAATCTCTGACAAAGAGCAGCTCT 195
QY      1844 TAAGTTCACAAAGTGGCTGGCTTGATTTGAATATCA-TTTCCTTTGCTATTTCCATTT 1902
Db      194 TAAGTTCACAAAGTGGCTGGCTTGATTTGAATATCA-TTTCCTTTGCTATTTCCATTT 135
QY      1903 ACATGAAAACCTTTGACTGTGAACTTGCATCTGTATTAATAATTAATCCCAATTAAG 1962
Db      134 ACATGAAAACCTTTGACTGTGAACTTGCATCTGTATTAATAATTAATCCCAATTAAG 76
QY      1963 GTTCTGTATGAAATGCTCTTTTATGCTTAAATTAATTAATTAATTAATTAATTAAT 2022
Db      75 GTTCTGTATGAAATGCTCTTTTATGCTTAAATTAATTAATTAATTAATTAATTAATTAAT 16
QY      2023 ATGGGATCCTAA 2034
Db      15 ATGGGATCCTAA 4

RESULT 11
BG963298      783 bp      mRNA      linear      EST 12-JUN-2001
LOCUS      602827716F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4982643 5',
DEFINITION      mRNA sequence.
ACCESSION      BG963298
VERSION      BG963298.1 GI:14350935
KEYWORDS      EST.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE      NIH-MGC http://mgi.nci.nih.gov/.
COMMENT      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished
CONTACT      Robert Strausberg, Ph.D.
E-MAIL      cgabbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10985 row: P column: 04
High quality sequence stop: 741.
Location/Qualifiers
1..783
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4982643"
/lab_host="DH10B (T1 phage-resistant)"
/clone_id="NCI_CGAP_Co24"
/note="Organ: colon; Vector: pCMV-SPORT6; Site: J. NCI;
Site_2: Salt; Cloned unidirectionally. Primer: Oligo dt.

```

```

BASE COUNT      141 a      236 c      226 g      180 t
ORIGIN
Query Match      18.9%; Score 385.2; DB 12; Length 783;
Best Local Similarity 77.2%; Pred. No. 1e-90;
Matches 545; Conservative 0; Mismatches 153; Indels 8; Gaps 6;

QY      128 ATGCTACGCTTAAAAAAGAAAGCTGCTGCTTCTGTTTCCAGATGGCATCTTG 187
Db      67 ATGCTACGCTTAAAAAAGAAAGCTGCTGCTTCTGTTTCCAGATGGCATCTTG 126
QY      188 GCTATTTCTTCCATGATGACACCAACATCAGCTCCCTGTATGAAGCAGACCC 247
Db      127 GCTCTTTCATTCATATGTCGTCGACAGACACTTCCAGAGGAGAGTCACAGAG- 185
QY      248 GAGCGATGACGCTGCTGCTGCTTCTGCTTCTGCGGCTCTGCTTCTTTTGGGGCAG 307
Db      186 --GCCGTCATGTGCTGCTGCTTCTGCTTCTGCGGCTCTGCTTCTTTTGGGGCAG 243
QY      308 CTTTGGGCGACACCCAGATGTTTCTTCTGATGAGCCCGCTGGCAGCTGGATG 367
Db      244 CTTTGGGCGACACCCAGATGTTTCTTCTGATGAGCCCGCTGGCAGCTGGATG 303
QY      368 ACCTTCAACGACGACCCCTGATGTCACATGCTGTGGGAGATGTATACGGGCC 427
Db      304 ACTTTCACGACGACACCCCTGATGTCACATGCTGTGGGAGATGTATACGGGCC 363
QY      428 GTCTTCTTGGGACATGAGCGCTTGTATGCTTATGAAACCTGCTCCCGGAGACAG 487
Db      364 GTCTTCTTGGGACATGAGCGCTTGTATGCTTATGAAACCTGCTCCCGGAGACAG 423
QY      488 TCCAGCCTTTCAGTGGAGAGAACACCGGCGCTGTCTTTCACCTGCTGTGACATC 547
Db      424 TCCAGCCTTTCAGTGGAGAGAACACCGGCGCTGTCTTTCACCTGCTGTGACATC 483
QY      548 ATCCCAAGATGAATATATCCCGGCGCTCTACATGAGCTCTGTGACATGACGCC 607
Db      484 TTCCCTGCGCCACGAGATGATGCTACCCAAAGCACTGCTGTGCGGTGACGAGCCC 543
QY      608 TTTGAGTGGTGGAGAGAGCGCTGCTCTTACAGCAGCAGTGGTGTGCA-GGAGGTGG 666
Db      544 TTTGATATGATGGAGAGAGCGCTCTCTACAGCTTGTGTGATCAAGAGGTGG 603
QY      667 -CTTCTCAACCTGACGCTCTTACCCGCTGTGAAAGACCCCTCTCAACCTG-CAT 724
Db      604 TATTCCTCAGCGTCGAGCGCTCTTATTCACAGTCTGCTCAACCTGCTCAC 663
QY      725 ATGCTGACCTGCTGCGGAGACCCCGGCGCTGTCTTCCCGAGAGACGACAAAGGA 784
Db      664 GTGCTGACCTGCTGCGGAGACCCCGGCGCTGTCTTCCCGAGAGACGACAAAGGA 722
QY      785 GATCTCATGATGA-CAGTCGATTTGATGGGAGCAGATGACAA 829
Db      723 GAATCTGTGTGACAGTCACTCTGCTGATGGGAGCAGATTTGAAA 768

RESULT 12
BG964671      695 bp      mRNA      linear      EST 12-JUN-2001
LOCUS      602831875F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4986258 5',
DEFINITION      mRNA sequence.
ACCESSION      BG964671
VERSION      BG964671.1 GI:14352308
KEYWORDS      EST.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE      NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)

```

JOURNAL
COMMENT

Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM10995 row: f column: 19
High quality sequence stop: 684.
Location/Qualifiers

FEATURES

source

1. .695
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4986258"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 118 a 216 c 195 g 166 t
ORIGIN

Query Match 18.0%; Score 368.6; DB 12; Length 695;
Best Local Similarity 74.9%; Pred. No. 2.6e-86;
Matches 501; Conservative 0; Mismatches 164; Indels 4; Gaps 3;

111 TCTTCCACTTGCACATGCTACTGCTTAATAAATGAAGCTCCTGCTGTTGTTGTT 170
19 TCTCTCTTCTCCGAGATGATGCTGTGAAGAAGAGAGCTGCTGATGTTCCGGGTT 78
171 CCCAGATGAGCCATCTTGGCTCTATCTTCCATGATGACGACACACATCAGTCCCGT 230
79 CCCAGATGAGGATGATGCTCTCTTCCATGATGCTGCTGCTGCTGCTGCTGCTGCT 138
231 CTATGAAGGACACCCGAGCGATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 290
139 --AGGAGAGAGTCCAGAGAGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 196
291 CTCTCTTCTGCGGAGCTTCTTGGGAGACCCAGATGTTTCTACTGATGAGACCCG 350
197 CCTC-TTGTGTGGAGAGCTTGTGGGAGACCCGAGATGTTTCTACTGATGAGACCCG 255
351 CCTGCACTGTGTGATGACCTTCAAGCAGACGACCGCTGATGCTGACATGCTGTGC 410
256 CCTGCACTGTGTGATGACCTTCAAGCAGACGACCGCTGATGCTGACATGCTGTGC 315
411 GGGATGATACGGGCGCTCTTGTGCGACATGAGCGTCTTGTGATGCTTACATGAGAC 470
316 GGGATGATACGGGCGCTCTTGTGCGACATGAGCGTCTTGTGATGCTTACATGAGAC 375
471 CTGTCCCGGAGAGACGACGCTTCTTGGGAGAGACGCGGCGCTGCTGTG 530
376 CAGGCGCCCGAAGACGACGCTTCTTCCAGTGGAGAGAGCGGCGCTGCTGTAG 435
531 CACCTGCTGTGATCATCCACAGATGAATCATCCCGGAGCTCATCTGACGAGCTCC 590
436 CGCTGTGTGATCTTCTTCCGCTCCAGATGATGATCACCAGACGCTGTGCTG 495
591 TGTGATCATCAACACCCCTTGTAGGTGTGAGAGAGGCTGCGCTCCTACAGCCAGCTG 650
496 TGTGATCATCAACACCCCTTGTAGGTGTGAGAGAGGCTGCGCTCCTACAGCCAGCTG 555
651 TGTGATCATCAACACCCCTTGTAGGTGTGAGAGAGGCTGCGCTCCTACAGCCAGCTG 710
556 TGTGATCATCAACACCCCTTGTAGGTGTGAGAGAGGCTGCGCTCCTACAGCCAGCTG 615
711 -CCCTCAACGTCATATGTCGACCTGCTCGGAGACCCCGGCGCTGCTGCTGCTGCTG 769

|||||
Db 616 CCCCTCAACCTGACAGCTGCTGACCTGCTCCAGAGACCCGCGCTGCTCCGATCCCGG 675
Oy 770 GAGGACACA 778
Db 676 GAGGACACA 684

RESULT 13

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM10998 row: c column: 08
High quality sequence stop: 784.
Location/Qualifiers

FEATURES

source

1. .852
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4987327"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 176 a 248 c 244 g 184 t
ORIGIN

Query Match 17.8%; Score 364.4; DB 12; Length 852;
Best Local Similarity 74.3%; Pred. No. 3.5e-85;
Matches 502; Conservative 0; Mismatches 166; Indels 8; Gaps 3;

116 CACTTCAACACATGCTACTGCTTAATAAATGAAGCTCCTGCTGTTGTTGTTGTT 175
125 CAGTGTATCAGATGATCTGTTGAAGAAGAGAGCTGATGATGTTCTGCTGCTCCAG 184
176 ATGGCATTCTGCTCTATCTTCTTCCATGATGACACCAACATCAGCTCTGTATG 235
185 GTCACTGTGTGATGCTCTTCAATCAATGTCGCGACAGACACCTTCCAG--AGG 241
236 AAGCAGACCCGAGAGGATGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 295
242 GAGGATGTCAGAGAGGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 301
296 -TTTGTGGGAGAGCTTCTTGGGAGACACCAAGATGTTTCTACCTGATGAGACCCGCTG 354
302 ATATGTGGAGAGCTTCTTGGGAGACACCCGAGATGTTTCTACCTGATGAGACCCGCTG 361
355 GCAGGTGTGATGATCTTCAAGCAGAGACCGCTGATGCTGACATGCTGCTGCGGGA 414

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 11:06:37 ; Search time 7289.74 Seconds

(Without alignments)
11465.210 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043

Sequence: 1 gaattccatgctgttggtga.....tgggacccctaaaaa 2043

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_da:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_ph:*
- 7: gb_pl:*
- 8: gb_pl:*
- 9: gb_pl:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_htg_hum:*
- 31: em_htg_in:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rtd:*
- 36: em_htg_man:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1915.2	93.7	2037	9	BC035282
2	1896	92.8	1992	9	AF280088
3	1867.4	91.4	2011	9	AK026635
4	1856.8	90.9	1979	6	BD127258
5	1856.8	90.9	1979	9	AK074746
6	1856.6	90.9	183228	9	AC010547
7	1856.6	90.9	206943	2	AC138848
8	1813	88.7	2063	6	AK310485
9	1784.8	87.4	2032	6	AR203335
10	1784.8	87.4	2032	6	BD134772
11	1784.8	87.4	2032	9	AF131235
12	1262	61.8	1333	9	AF149783
13	763.6	37.4	877	6	BD124832
14	763.6	37.4	877	6	BD126950
15	755	37.0	2201	10	AF109155
16	752	36.8	232745	2	AC130174
17	743.2	36.4	1926	6	BD134773
18	743.2	36.4	1926	10	AF131236
19	708.2	34.7	262679	2	AC110314
20	480.2	23.5	517	6	AK381256
21	412.8	20.2	1647	6	AK327330
22	412.8	20.2	2544	9	AF280086
23	412.8	20.2	3786	9	AF219991
24	412.8	20.2	71503	9	AC009163
25	412.8	20.2	157358	9	AC009163
26	412.8	20.2	208185	2	AC009105
27	398.4	19.5	1462	9	AF176839
28	398.4	19.5	2170	9	AF176838
29	398.4	19.5	3278	9	AF246718
30	398.4	19.5	194832	9	AC025287
31	390.4	19.1	229007	2	AC095664
32	390.4	19.1	245248	2	AC117869
33	387.2	19.0	1740	10	AF176841
34	387.2	19.0	1989	10	AF176840
35	387.2	19.0	159170	2	AC115914
36	387.2	19.0	209198	2	AC127315
37	376.2	18.4	548	6	BD125617
38	294.4	14.4	389	6	AK381326
39	144.8	7.1	6961	9	AB017915
40	144.8	7.1	178253	9	AC073370
41	144.8	7.1	196467	9	AC022392
42	141.6	6.9	2156	6	AR071396
43	141.6	6.9	2156	6	E14937
44	141.6	6.9	2731	9	AB012192
45	136.8	6.7	200269	9	AC087442

ALIGNMENTS

RESULT 1
BC035282
LOCUS
DEFINITION BC035282 2037 bp. mRNA. linear. PRI 23-SEP-2002
Homo sapiens. Similar to carboxylate (N-acetylglucosamine 6-o)
sulfotransferase 4, clone MGC:34456 IMAGE:5179826, mRNA, complete
cds.
ACCESSION BC035282
VERSION BC035282.1 GI:23273964
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2037)
AUTHORS Strausberg,R.

TITLE
JOURNAL

Direct Submission
Submitted (31-Jul-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Huljk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

FEATURES

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
Series: IRAK Plate: 50 Row: a Column: 7
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 5031734.

Source

CDS

Location/Qualifiers
1..2037
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:34456 IMAGE:5179826"
/tissue_type="Brain, Lung, Testis, adult, pooled whole"
/clone_1lb="NIH_MGC_115"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
126..1238
/codon_start=1
/product="Similar to carbohydrate (N-acetyl)glucosamine
6-O) sulfotransferase 4"
/protein_id="AAH35282.1"
/db_xref="GI:23273965"
/translation="MAILALFFHMYSHNISLSMKRQPRMHLVLSWRSSSFVGO
LEGHDPVYLEMPAHVWMTKOSTAMHVAVDLRAVFLCQMSVDAWMEGPR
ROSLFQWNSALCSAPACDIIRODEIIPRAHRLCSQVPEVVERKRSYHVL
KEVERFNQSLPLKDPSLNHIYHVAHVPRAVRSKRITGDLIDRIWQDHO
KLRKEDQPYVMQVLCQSOLELYKTQSLPKALOERYLIVREDLARAVACTSRMTE
FVGLFELPHLQTMVHNITRGKMGMDHAFETNADLNVLSOAWRMSPYKRVSLQKAC
GDANMLLGRVHRSRQEQRNLLDLILSTWVPEQIH"

BASE COUNT

507 a 546 c 481 g 503 t

Query Match

Best Local Similarity 93.7%; Score 1915.2; DB 9; Length 2037;
Matches 1978; Conservative 0; Mismatches 13; Indels 5; Gaps 5;

QY 51 GGGCATGGCGGGCTAGAGAGCTCTCAAAAGAGAGAGGAAGCCCAAGCCCAAG 110
DB 1 GGGCATGGCGGGCTAGAGAGCTCTCAAAAGAGAGAGGAAGCCCAAGCCCAAG 60
QY 111 TCTTCACCTTACAGCAATGCTACTGCTCTCTCTCTCTCTCTCTCTCTCTCT 170
DB 61 TCTTCACCTTACAGCAATGCTACTGCTCTCTCTCTCTCTCTCTCTCTCTCT 120
QY 171 CCCAGATGGCATCTTGGCTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 230
DB 121 CCCAGATGGCATCTTGGCTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
QY 231 CTATGAAGCAGACCCGAGCGCATGACAGTGTCTGTCTCTCTCTCTCTCTCT 290
DB 181 CTATGAAGCAGACCCGAGCGCATGACAGTGTCTGTCTCTCTCTCTCTCTCT 240
QY 291 CTCTCTTGTGGGAGCTTTTGGGACAGCCAGATTTTCTACCTGATGAGGCCG 350

DB 241 CTCTCTTGTGGGAGCTTTTGGGACAGCCAGCATGTTTCTACCTGATGAGCCCG 300
QY 351 CCGGACAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 410
DB 301 CCGGACAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
QY 411 GGGATCTGATACGGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 470
DB 361 GGGATCTGATACGGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
QY 471 CCGGACAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 530
DB 421 CCGGACAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480
QY 531 CACCTCTCTGTATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 590
DB 481 CACCTCTCTGTATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 540
QY 591 TGTGCAATCAACAGCCCTTTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 650
DB 541 TGTGCAATCAACAGCCCTTTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
QY 651 TGTGCAATCAACAGCCCTTTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 710
DB 601 TGTGCAATCAACAGCCCTTTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 660
QY 711 CCTCAACCTGCATATCATCATCATCATCATCATCATCATCATCATCATCATCAT 770
DB 661 CCTCAACCTGCATATCATCATCATCATCATCATCATCATCATCATCATCATCAT 720
QY 771 AACGCAAAAGGAGATCTCATGATGATGATGATGATGATGATGATGATGATGATG 830
DB 721 AACGCAAAAGGAGATCTCATGATGATGATGATGATGATGATGATGATGATGATG 780
QY 831 AACTCAAGAAG 890
DB 781 AACTCAAGAAG 840
QY 891 AGATCTCAAGACCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 950
DB 841 AGATCTCAAGACCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 900
QY 951 GCTATGAG 1010
DB 901 GCTATGAG 960
QY 1011 GATGGAATCTTGGCCCATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1070
DB 961 GATGGAATCTTGGCCCATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
QY 1071 TGGGATGACAGCGCTTTCACACAAATGCGAGGAGATGCCCTTAATGCTCCAG 1130
DB 1021 TGGGATGACAGCGCTTTCACACAAATGCGAGGAGATGCCCTTAATGCTCCAG 1080
QY 1131 GCTGATGAG 1190
DB 1081 GCTGATGAG 1140
QY 1191 ATTTGCTGAG 1250
DB 1141 ATTTGCTGAG 1200
QY 1251 TTTCTGTCTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1310
DB 1201 TTTCTGTCTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1311 CCACCTGAG 1370
DB 1261 CCACCTGAG 1320
QY 1371 GAGGCTTAATCAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1430
DB 1321 GAGGCTTAATCAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1380

```

QY 1431 AACGAGAGACTTTTGTGTCATGCTGTGTCTAGAAAAACAGACTGGGAACTTATGT 1490
    |||
Db 1381 AACGAGAGAGACTTTTGTGTCATGCTGTGTCTAGAAAAACAGACTGGGAACTTATGT 1440
QY 1491 GAGCAGACATCCACCAAGTGAACAGGGTATTGCTCTCTCTCTCTCTCTCTCTCTCT 1550
    |||
Db 1441 GAGCAGACATCCACCAAGTGAACAGGGTATTGCTCTCTCTCTCTCTCTCTCTCTCT 1500
QY 1551 TCTGGGACAGCTTCAGAGACTTGTGCTGCGGAGGCGCTATTAGACAGACAGTATCAG 1610
    |||
Db 1501 TCTGGGACAGCTTCAGAGACTTGTGCTGCGGAGGCGCTATTAGACAGACAGTATCAG 1560
QY 1611 TGGAAATGATCCATAAATCCCTGTCACATCTGCCAATGGGAATGATCTTTAC 1670
    |||
Db 1561 TGGAAATGATCCATAAATCCCTGTCACATCTGCCAATGGGAATGATCTTTAC 1620
QY 1671 CAAGAGCTCACCAGCATTTTCCACAGAGATGCGAAATTCGAGCCCTTGGAGTCCCAAT 1730
    |||
Db 1621 CAAGAGCTCACCAGCATTTTCCACAGAGATGCGAAATTCGAGCCCTTGGAGTCCCAAT 1680
QY 1731 GGGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1789
    |||
Db 1681 -GGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1739
QY 1790 AGCTATCGGTATCAAGAAATATGAAACAAATCTCTG-ACAAAAGAGAGAGAGAGAGAG 1848
    |||
Db 1740 AGCTATCGGTATCAAGAAATATGAAACAAATCTCTGACACAAAGAGAGAGAGAGAGAG 1799
QY 1849 TCACAGAGTGGCTGGGCTTGATTTGAATATCA-TTCCCTTTCATTTTCCATTCATA 1907
    |||
Db 1800 TCACAGAGTGGCTGGGCTTGATTTGAATATCACTTCCCTTTCATTTTCCATTCATA 1859
QY 1908 GAAACTTTGACCTGGGAACTTGCATCTGTATTAATTAATTAATTAATTAATTAATTAAT 1967
    |||
Db 1860 GAAACTTTGACCTGGGAACTTGCATCTGTATTAATTAATTAATTAATTAATTAATTAAT 1918
QY 1968 GTTTAGAAATGCTCTTTTATGCTTTATGCTTTATTAATTAATTAATTAATTAATTAAT 2027
    |||
Db 1919 GTTTAGAAATGCTCTTTTATGCTTTATGCTTTATTAATTAATTAATTAATTAATTAAT 1978
QY 2028 ATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2043
    |||
Db 1979 ATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1994

RESULT 2
AF280088. 1992 bp mRNA linear PRI 20-FEB-2001
LOCUS Homo sapiens I-selectin ligand sulfotransferase GST-3 mRNA,
DEFINITION complete cds.
ACCESSION AF280088
VERSION AF280088.1 GI:12060807
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1992)
AUTHORS Hemmerich,S., Lee,J.K., Bhakta,S., Bistrup,A., Ruddle,N.R. and
Rosen,S.D.
TITLE Chromosomal localization and genomic organization for the
JOURNAL galactose/N-acetylgalactosamine/N-acetylglucosamine
MEDLINE 6-O-sulfotransferase gene family
PUBMED glycobiology 11 (1), 75-87 (2001)
21096027
11181564
REFERENCE 2 (bases 1 to 1992)
AUTHORS Hemmerich,S., Bhakta,S., Lee,J.-K., Bistrup,A., Ruddle,N.R. and
Rosen,S.D.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience,
3401 Hillview Ave., Palo Alto, CA 94304, USA

```

```

FEATURES
    source
        Location/Qualifiers
            1..1992
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /chromosome="16"
                /map="16q23.1-q23.2"
                /cell_type="high endothelial"
                /tissue_type="tonsil"
            1..1992
                /gene="GST3"
                /gene="GST3"
                /note="similar to Homo sapiens sequence encoded by GenBank
                Accession Number AF131235"
            /codon_start=1
            /product="I-selectin ligand sulfotransferase GST-3"
            /protein_id="AA048246.1"
            /db_xref="GI:12060808"
            /translation="MLPKMKLLFLVSOMALILFPHYSHNISLSMKAPERH
            VIVLSMRSSSFVGLFGHPDVFYLMEMAHVMTFROSTAMLMHAYRDLIRAY
            LCDMSYFDATMEGRPRLQSLFPMENSRALCSRPACDIITODEITIRARSLRCSOP
            FEVEKACRSYSHVLEKVEFNQSLYPLKPSLNHLVLRDPRALFRERERK
            GDLMIDSRIVMGHQLKKEDQPYVMQVYICQSOLEIYTIQSLPRALQERYLVRY
            EDLARPVQTSRYEFVGLFPLHQTWVHNITRKGMDHAFHTNARDALNVSOAM
            RMSLPYEXSRILQKACGDAMNLLGVRHVRSEQGRNLLDLSTWTVPEQIH"
        BASE COUNT
            482 a 540 c 489 g 481 t
            ORIGIN
                Query Match 92.8%; Score 1896; DB 9; Length 1992;
                Best Local Similarity 99.2%; Pred. No. 0;
                Matches 1957; Conservative 0; Mismatches 10; Indels 5; Gaps 5;

```

```

QY 607 CTTGAGTGTGTGAGAGAGGCTGCCCTCTACAGCCAGTGTGTCTCAGAGAGTGGC 666
DB 601 CTTTGAAGGTGTGAGAGAGGCTGCCCTCTACAGCCAGTGTGTCTCAGAGAGTGGC 660
QY 667 CTTCTTCAACCTGAGTGTCTTACCCGCTGTGGAAGAGCCCTCTCAGAGAGTGTAT 726
DB 661 CTTCTTCAACCTGAGTGTCTTACCCGCTGTGGAAGAGCCCTCTCAGAGAGTGTAT 720
QY 727 GGTGACCTGTGTGAGAGAGGCTGCCCTCTACAGCCAGTGTGTCTCAGAGAGTGGC 786
DB 721 GGTGACCTGTGTGAGAGAGGCTGCCCTCTACAGCCAGTGTGTCTCAGAGAGTGGC 780
QY 787 TGTGATGATGAGAGAGGCTGTGTGTGAGAGAGTGTGTGTGTGTGTGTGTGTGTGT 846
DB 781 TGTGATGATGAGAGAGGCTGTGTGTGAGAGAGTGTGTGTGTGTGTGTGTGTGTGT 840
QY 847 CCAACCCCTACTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 906
DB 841 CCAACCCCTACTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 900
QY 907 CCAATCTCTTGTGAGAGAGGCTGCCCTCTACAGCCAGTGTGTGTGTGTGTGTGTGT 966
DB 901 CCAATCTCTTGTGAGAGAGGCTGCCCTCTACAGCCAGTGTGTGTGTGTGTGTGTGT 960
QY 967 TGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1026
DB 961 TGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1020
QY 1027 CCAATCTCTTGTGAGAGAGGCTGCCCTCTACAGCCAGTGTGTGTGTGTGTGTGTGT 1086
DB 1021 CCAATCTCTTGTGAGAGAGGCTGCCCTCTACAGCCAGTGTGTGTGTGTGTGTGTGT 1080
QY 1087 CCAACAAATGAGAGAGGCTGCCCTCTACAGCCAGTGTGTGTGTGTGTGTGTGTGT 1146
DB 1081 CCAACAAATGAGAGAGGCTGCCCTCTACAGCCAGTGTGTGTGTGTGTGTGTGTGT 1140
QY 1147 TGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1206
DB 1141 TGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1200
QY 1207 CCAAGTCAATGTGAAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1266
DB 1201 CCAAGTCAATGTGAAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1260
QY 1267 TGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1326
DB 1261 TGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1320
QY 1327 TGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1386
DB 1321 TGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1380
QY 1387 TGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1446
DB 1381 TGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1440
QY 1447 TGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1506
DB 1441 TGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1500
QY 1507 TGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1566
DB 1501 TGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1560
QY 1567 TGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1626
DB 1561 TGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1620
QY 1627 TGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1686
DB 1621 TGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1680
QY 1687 TGTGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1746

```

```

DB 1681 ATTTTCCACAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1739
QY 1747 AGTGGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1805
DB 1740 AGTGGAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1799
QY 1806 AATATGAACAAATCTGTG-AACAAAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1864
DB 1800 AATATGAACAAATCTGTG-AACAAAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1859
QY 1865 CTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1923
DB 1860 CTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1919
QY 1924 GAACCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1975
DB 1920 GAAGC-TGCCATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1970

RESULT 3
AK026635
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
misc_feature
BASE COUNT
ORIGIN
Query Match
Best Local Similarity

```

Matches 1932; Conservative 0; Mismatches 16; Indels 5; Gaps 5;				
QY	94	AAGCCAAAGCCAAAGGCTTCCACTTCCAGACATGCTACTGCTTAAAAAATGAACCT	153	
Db	38	AAGCCGCTCTTGAAGGCTTCCACTTCCAGACATGCTACTGCTTAAAAAATGAACCT	97	
QY	154	CCCTGCTTTTGGTTTCCAGATGAGCCATTTGGCTTATTTCTTCCATGTACAGCA	213	
Db	98	CCCTGCTTTTGGTTTCCAGATGAGCCATTTGGCTTATTTCTTCCATGTACAGCA	157	
QY	214	CAACATCAGCTCCCTGTCTATGAAGGACAGCCGAGCGCATGCATGCTGTTCTGTC	273	
Db	158	CAACATCAGCTCCCTGTCTATGAAGGACAGCCGAGCGCATGCATGCTGTTCTGTC	217	
QY	274	TTCTGAGGCTGTGGCTTCTTTTGGGGGAGGTTTGGGAGCCCAAGATTTT	333	
Db	218	TTCTGAGGCTGTGGCTTCTTTTGGGGGAGGTTTGGGAGCCCAAGATTTT	277	
QY	334	CTACCTGATGAGCCCGGCTGGACGTGTGATGACCTTCAAGCAGACCGCCTGGAT	393	
Db	278	CTACCTGATGAGCCCGGCTGGACGTGTGATGACCTTCAAGCAGACCGCCTGGAT	337	
QY	394	GCTGACATGCTGTGGGGGATCTGATACGGGCGCTCTTGTGCGCATGAGCGTCTT	453	
Db	338	GCTGACATGCTGTGGGGGATCTGATACGGGCGCTCTTGTGCGCATGAGCGTCTT	397	
QY	454	TGATGCCATACATGAGAACCTGGTCCCGGAGACAGCCGCTTTGATGGGAGAAAG	513	
Db	398	TGATGCCATACATGAGAACCTGGTCCCGGAGAGCCGCTTTGATGGGAGAAAG	457	
QY	514	CCGGGCGCTGTGTTCTGACACCTGCTGTGATCATTCACCAAGATGAATCATCCCG	573	
Db	458	CCGGGCGCTGTGTTCTGACACCTGCTGTGATCATTCACCAAGATGAATCATCCCG	517	
QY	574	GGCTCATGCGAGGCTCCCTGTGACATCAACACAGCCCTTTAGTGTGGGAAAGGCTGCC	633	
Db	518	GGCTCATGCGAGGCTCCCTGTGACATCAACACAGCCCTTTAGTGTGGGAAAGGCTGCC	577	
QY	634	CTCTACAGCAGCAGTGTGCTCAAGAGAGTGGCTTCTTCAACGCGAGTCCCTTACC	693	
Db	578	CTCTACAGCAGCAGTGTGCTCAAGAGAGTGGCTTCTTCAACGCGAGTCCCTTACC	637	
QY	694	GCTGCTGAAGACCCCTCCCTCAACCTGCAATATGCTGACCTGTGTCGGGACCCCGGGC	753	
Db	638	GCTGCTGAAGACCCCTCCCTCAACCTGCAATATGCTGACCTGTGTCGGGACCCCGGGC	697	
QY	754	CGTGTCCGTTCCCGAGAGGACCAAGAGAGATCTCATGATGTGACATGCTGAT	813	
Db	698	CGTGTCCGTTCCCGAGAGGACCAAGAGAGATCTCATGATGTGACATGCTGAT	757	
QY	814	GGGCGAGCATGAGCAAAATCAAGAGAGAGACCAACCTACTGTGATGAGAGT	873	
Db	758	GGGCGAGCATGAGCAAAATCAAGAGAGAGACCAACCTACTGTGATGAGAGT	817	
QY	874	CTGCCAAAGCAGCTGAGATCTACAGACCATCCAGTCTTGCCCAAGGCCCTGACAGA	933	
Db	818	CTGCCAAAGCAGCTGAGATCTACAGACCATCCAGTCTTGCCCAAGGCCCTGACAGA	877	
QY	934	ACGCTACCTGCTTGGCGCTATGAGAGCTGGCTGAGACCCCTGTGGGCCAGACTTCCG	993	
Db	878	ACGCTACCTGCTTGGCGCTATGAGAGCTGGCTGAGACCCCTGTGGGCCAGACTTCCG	937	
QY	994	AATGTATGAATTCGTGGATTTGAATTCCTCCCATTTTGAAGACTGGTGCATTAACAT	1053	
Db	938	AATGTATGAATTCGTGGATTTGAATTCCTCCCATTTTGAAGACTGGTGCATTAACAT	997	
QY	1054	CACCCGAGGCAAGGCGATGGTGAACACGCTTTTCCACAACAATGCCAGGATGCCCTTAA	1113	
Db	998	CACCCGAGGCAAGGCGATGGTGAACACGCTTTTCCACAACAATGCCAGGATGCCCTTAA	1057	
QY	1114	TGCTCCAGGCTTGGGCGTGGCTTCTTCCCTATGAAGAAAGTTTCTGCATCTTCAAGAAC	1173	
Db	1058	TGCTCCAGGCTTGGGCGTGGCTTCTTCCCTATGAAGAAAGTTTCTGCATCTTCAAGAAC	1117	

QY	1174	CTGTGGCGATGCGATGATTTGCTGGGCTTACCGCCACGCGAGATCTGAACAAGACAGAG	1233	
Db	1118	CTGTGGCGATGCGATGATTTGCTGGGCTTACCGCCACGCTGATCTGAACAAGACAGAG	1177	
QY	1234	AAACCTGTGTGGATCTTGTGTACCTGTGAGCTGTCCCTGTGAGCAAAATCCATAAGAGG	1293	
Db	1178	AAACCTGTGTGGATCTTGTGTACCTGTGAGCTGTCCCTGTGAGCAAAATCCATAAGAGG	1237	
QY	1294	TTGAGAGGCTTGTGCTGCACCTGTGTGACCTTCACTTCTGTGATGCTTCTGA	1353	
Db	1238	TTGAGAGGCTTGTGCTGCACCTGTGTGACCTTCACTTCTGTGATGCTTCTGA	1297	
QY	1354	GCCCTGCTACATCTGAGCCTTACACATATGCTGTGGGATACACTGAGCTGTGAGT	1413	
Db	1298	GCCCTGCTACATCTGAGCCTTACACATATGCTGTGGGATACACTGAGCTGTGAGT	1357	
QY	1414	TGTGTCCACAGTGTGCTCAAGCAGAAAGACTTTTGTCCATGCTTGTGTGAGAAAGAG	1473	
Db	1358	TGTGTCCACAGTGTGCTCAAGCAGAAAGACTTTTGTCCATGCTTGTGTGAGAAAGAG	1417	
QY	1474	ACTGGGGAACCTTATGTGAGCAGCATTCACACAGTGAAGAGGATTTGCTTCTTC	1533	
Db	1418	ACTGGGGAACCTTATGTGAGCAGCATTCACACAGTGAAGAGGATTTGCTTCTTC	1477	
QY	1534	TTTCTGTATCTTCTGCTGGGAGACTTTCAGACATTTGTGGCTGGAGGCTTATTA	1593	
Db	1478	TTTCTGTATCTTCTGCTGGGAGACTTTCAGACATTTGTGGCTGGAGGCTTATTA	1537	
QY	1594	GCACGACACAGTATCAGTGAATTTGATTCATTAACCTCCCTGTCCACATCTTGCCCAATG	1653	
Db	1538	GCACGACACAGTATCAGTGAATTTGATTCATTAACCTCCCTGTCCACATCTTGCCCAATG	1597	
QY	1654	GGGAATGATCTTTTACCAAGAGCTCACACAGATTTTCCACAGAGATGCGAATTTCTGAG	1713	
Db	1598	GGGAATGATCTTTTACCAAGAGCTCACACAGATTTTCCACAGAGATGCGAATTTCTGAG	1657	
QY	1714	CCCTTGGAGTCCCAATGGGATTTCAAGAGAGGATGGGAAACAAAGTGGATGCTTACTT	1773	
Db	1658	CCCTTGGAGTCCCAATGGGATTTCAAGAGAGGATGGGAAACAAAGTGGATGCTTACTT	1716	
QY	1774	ATGAGCTTGACCAT-ACAGCTATGCTGTAATCAGAAATATGAAACAAATCTCTG-ACAAA	1831	
Db	1717	ATGAGCTTGACCATCAGCTATGCTGTAATCAGAAATATGAAACAAATCTCTGACAAA	1776	
QY	1832	AGAGCAAGCTCTTAAGTTCACAGAGGCTGGGCTGTTGAATTAATCA-TTCCCTTTG	1890	
Db	1777	AGAGCAAGCTCTTAAGTTCACAGAGGCTGGGCTGCTGAATTAATTAATCACTCCCTCTG	1836	
QY	1891	CATTTCCCATTTACATAGAAAACCTTGAACCTGTGAACCTGGCATCTGTAACTAATAA	1950	
Db	1837	CATTTCCCATTTACATAGAAAACCTTGAACCTGTGAACCTGGCATCTGTAACTAATAA	1895	
QY	1951	TTCCCAAAATTAAGGTTCTGTTAGAAATGTCCTTTTATGCTTCTTAATTAATAGCAGTAA	2010	
Db	1896	TTCCCAAAATTAAGGTTCTGTTAGAAATGTCCTTTTATGCTTCTTAATTAATAGCAGTAA	1955	
QY	2011	ATGTTCAATTTTATGAGATCCCTAATAAAAAA 2043		
Db	1956	ATGTTCAATTTTATGAGATCCCTAATAAAAAA 1988		
RESULT 4				
BD127258				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
Homo sapiens (human)				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				

REFERENCE	1 (bases 1 to 1979)
AUTHORS	Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y., Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and Koga,H.
TITLE	Primer for synthesizing full-length cDNA and use thereof
JOURNAL	Patent: JP 200217375-A 2689 22-JAN-2002; HELIX RESEARCH INSTITUTE
COMMENT	OS Homo sapiens (human) PN JP 2002017375-A/2689 PD 22-JAN-2002 PE 07-JUL-2000 JP 2000253172 PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO PI ISHII PI YUKI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI PI SHINICHI KOJIMA, PI TETSUJI OTSUKI, HISASHI KOGA
FEATURES	FT CDS Location/Qualifiers 1..1979 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"
BASE COUNT	451 a 542 c 474 g 512 t
ORIGIN	
Query Match	90.98; Score 1856.8; DB 6; Length 1979; Best Local Similarity 98.98; Pred. No. 0; Matches 1922; Conservative 0; Mismatches 17; Indels 5; Gaps 5;
94	AAAGCCAGACAGTGTCTCCATTCAGACATGCTACTGCTTAAATGAAGCT 153
35	AAAGCCGCTTGAAGAGTCTCCATTCAGACATGCTACTGCTTAAATGAAGCT 94
154	CGTGTGTTTGGTGTCCAGATGAGCATCTTGGCTATTTCTTCACATGTACAGCA 213
95	CGTGTGTTTGGTGTCCAGATGAGCATCTTGGCTATTTCTTCACATGTACAGCA 154
214	CAACATGATGAGCCCGCTGAGCATGATGATGATGATGATGATGATGATGATGAT 273
155	CAACATGATGAGCCCGCTGAGCATGATGATGATGATGATGATGATGATGATGAT 214
274	TTCTGAGGCTGAGCTGCTTCTTTTGGGGGAGCTTTTGGGAGCAGCCAGATGAT 333
215	TTCTGAGGCTGAGCTGCTTCTTTTGGGGGAGCTTTTGGGAGCAGCCAGATGAT 274
334	CTACCTGATGAGCCCGCTGAGCATGATGATGATGATGATGATGATGATGATGAT 393
275	CTACCTGATGAGCCCGCTGAGCATGATGATGATGATGATGATGATGATGATGAT 334
394	GCTGACATGAGCTGCTGAGCATGATGATGATGATGATGATGATGATGATGATGAT 453
335	GCTGACATGAGCTGCTGAGCATGATGATGATGATGATGATGATGATGATGATGAT 394
454	TGATGCTATGATGAGCTGCTGAGCATGATGATGATGATGATGATGATGATGATGAT 513
395	TGATGCTATGATGAGCTGCTGAGCATGATGATGATGATGATGATGATGATGATGAT 454
514	CGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 573
455	CGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 514
574	GGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 633
515	GGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 574
634	CTCTACAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 693
575	CTCTACAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 634
694	GCTGCTGAGAGAGCCCTCCCTCAACCTGATGATGATGATGATGATGATGATGATGAT 753
635	GCTGCTGAGAGAGCCCTCCCTCAACCTGATGATGATGATGATGATGATGATGATGAT 694
754	CGTGTCCGTTCCCGAGACGACCAAGAGAGATCTCATGATTTGACAGTCCATTTGAT 813
695	CGTGTCCGTTCCCGAGAGACGACCAAGAGAGATCTCATGATTTGACAGTCCATTTGAT 754
814	GGGGGAGCATGACCAAACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 873
755	GGGGGAGCATGACCAAACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 814
874	CTGCCAAGCCAGCTGAGATCTACAGAGACATGATGATGATGATGATGATGATGATGAT 933
815	CTGCCAAGCCAGCTGAGATCTACAGAGACATGATGATGATGATGATGATGATGATGAT 874
934	ACGCTACCTGCTTGTGCTATGAGACCTGAGCTGAGCCCTGAGCCAGACTTCCG 993
875	ACGCTACCTGCTTGTGCTATGAGACCTGAGCTGAGCCCTGAGCCAGACTTCCG 934
994	AATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1053
935	AATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 994
1054	CACCCGAGCAAGGAGATGGGTGACCAAGCTTCCACACAAATGCCAGGATGCCCTTA 1113
995	CACCCGAGCAAGGAGATGGGTGACCAAGCTTCCACACAAATGCCAGGATGCCCTTA 1054
1114	TGCTCCAGAGCTTGGGCTGCTGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 1173
1055	TGCTCCAGAGCTTGGGCTGCTGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 1114
1174	CTGTGGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1233
1115	CTGTGGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1174
1234	AAACTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1293
1175	AAACTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1234
1294	TTGAGAGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1353
1235	TTGAGAGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1294
1354	GCCTGCTACATCTGAGGCTTAACTATGATGATGATGATGATGATGATGATGATGAT 1413
1295	GCCTGCTACATCTGAGGCTTAACTATGATGATGATGATGATGATGATGATGATGAT 1354
1414	TGTGTCCACAGCTGCTCAAGCAGAGACTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1473
1355	TGTGTCCACAGCTGCTCAAGCAGAGACTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1414
1474	ACTGGGAACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1533
1415	ACTGGGAACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1474
1534	TTTTCTTGAATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1593
1475	TTTTCTTGAATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1534
1594	GCAAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1653
1535	GCAAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1594
1654	GGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1713
1595	GGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1654
1714	CCCTTGAGTCCCATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1773
1655	CCCTTGAGTCCCATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1713


```

QY 994 AATGATGAATTCGTGGGATGGAATTCCTTCCCATCTTCAGACCTGGGTGATACAT 1053
DB 935 AATGATGAATTCGTGGGATGGAATTCCTTCCCATCTTCAGACCTGGGTGATACAT 994
QY 1054 CACCCGAGGAGGAGGATGGGTGATCCACAGCTTCCACACAAATCCAGGATGCCCTAA 1113
DB 995 CACCCGAGGAGGAGGATGGGTGATCCACAGCTTCCACACAAATCCAGGATGCCCTAA 1054
QY 1114 TGTCTCCAGAGCTTGGGCGGTCTTGGCCCTATGAAAGATTTCTGATGAGAAAGC 1173
DB 1055 TGTCTCCAGAGCTTGGGCGGTCTTGGCCCTATGAAAGATTTCTGATGAGAAAGC 1114
QY 1174 CTGTGGCGATGCCATGAATTTCTGGGCTACCGCCACGTAGATCTGACACAGACAG 1233
DB 1115 CTGTGGCGATGCCATGAATTTCTGGGCTACCGCCACGTAGATCTGACACAGACAG 1174
QY 1234 AATCGTTGCTGGATCTTCTGTCTACCTGAGTGCCTGAGCAATCCACTAAGAGG 1293
DB 1175 AATCGTTGCTGGATCTTCTGTCTACCTGAGTGCCTGAGCAATCCACTAAGAGG 1234
QY 1294 TTGAGAGGCTTGTGCTGCCACCTGTGTGACGCTCAGTCACTTCTGTAATGCTTGA 1353
DB 1235 TTGAGAGGCTTGTGCTGCCACCTGTGTGACGCTCAGTCACTTCTGTAATGCTTGA 1294
QY 1354 GCCTTGCCTACATCTCTGAGCCTTAATCTATGTCGTGGGTATCAGACTGAGTGTAGT 1413
DB 1295 GCCTTGCCTACATCTCTGAGCCTTAATCTATGTCGTGGGTATCAGACTGAGTGTAGT 1354
QY 1414 TGTGTCCACAGCTGTCTCAGCAGAGAGGACTTGTGTGCATGCTGTGTAGAAACAG 1473
DB 1355 TGTGTCCACAGCTGTCTCAGCAGAGAGGACTTGTGTGCATGCTGTGTAGAAACAG 1414
QY 1474 ACTGGGAACTTATGTGAGCAGACATCCCAAGTGAAGAGGATTTGCTCTCTTC 1533
DB 1415 ACTGGGAACTTATGTGAGCAGACATCCCAAGTGAAGAGGATTTGCTCTCTTC 1474
QY 1534 TTTTCTGATCTCTGTCTGGGAGAGCTTCAGAGACTTTGGCCCGGAGGCTTAA 1593
DB 1475 TTTTCTGATCTCTGTCTGGGAGAGCTTCAGAGACTTTGGCCCGGAGGCTTAA 1534
QY 1594 GCACGACACAGTATCAGTGAATTCATTAACCTCCTGTCCACATCTTCCCAATG 1653
DB 1535 GCACGACACAGTATCAGTGAATTCATTAACCTCCTGTCCACATCTTCCCAATG 1594
QY 1654 GGAATGAGATCTTTCACCAAGAGCTCAGCAGATTTCCACAGAGATGCGAATTTCTGAG 1713
DB 1595 GGAATGAGATCTTTCACCAAGAGCTCAGCAGATTTCCACAGAGATGCGAATTTCTGAG 1654
QY 1714 CCCTTGGAGTTCCCAATGGATTCAGGAAGAGTGGGAACAAGTTGATGCTTAA 1773
DB 1655 CCCTTGGAGTTCCCAATGGATTCAGGAAGAGTGGGAACAAGTTGATGCTTAA 1713
QY 1774 ATGAGCTTGACAT-ACAGCTATGCTGATCAGAAATATGAAACAAATCTCTGACAAA 1831
DB 1714 ATGAGCTTGACATCAGCTATGCTGATCAGAAATATGAAACAAATCTCTGACAAA 1773
QY 1832 AGGCAAGCTTAAATTCACAGGTGCTGGGCTGATTTGAATATCA-TTTCCTTTG 1890
DB 1774 AGGCAAGCTTAAATTCACAGGTGCTGGGCTGATTTGAATATCA-TTTCCTTTG 1833
QY 1891 CATTTTCCATACATAGAAATCTTGACCTGGAACCTGCCATGCTTAATATCTAA 1950
DB 1834 CATTTTCCATACATAGAAATCTTGACCTGGAACCTGCCATGCTTAATATCTAA 1892
QY 1951 TTCCCAAAATAGGTTCTGTTTAAATGTCCTTTTATGCTTCTTAATATTAAGCAGTAA 2010
DB 1893 TTCCCAAAATAGGTTCTGTTTAAATGTCCTTTTATGCTTCTTAATATTAAGCAGTAA 1952
QY 2011 ATGTTCATTTTATGAGATCTTAA 2034
DB 1953 ATGTTCATTTTATGAGATCTTAA 1976

```

```

RESULT 6
AC010547
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

AC010547 183228 bp DNA linear PRI 29-SEP-2001
Homo sapiens chromosome 16 clone RP11-510M2, complete sequence.

AC010547 GI:15808510
HTG.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 183228)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
2 (bases 1 to 183228)
DOE Joint Genome Institute.
Direct Submission
Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 183228)
DOE Joint Genome Institute.
Direct Submission
Submitted (29-SEP-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Sep 29, 2001 this sequence version replaced g1:14589436.

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

FEATURES
source
1. 183228
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-510M2"
BASE COUNT 51901 a 42098 c 40026 g 49203 t
ORIGIN

Query Match 90.9%; Score 1856.6; DB 9; Length 183228;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1920; Conservative 0; Mismatches 14; Indels 5; Gaps 5;

QY 107 AAGGTCTTCACCTGACACAAATGCTACGCTCCCAAAATGAAGTCCGCTGTTTCG 166
DB 142584 AAGGTCTTCACCTGACACAAATGCTACGCTCCCAAAATGAAGTCCGCTGTTTCG 142643

QY 167 GTTCCAGATGCGCATCTTGCTCTATTTCTTCACATGTACAGCCACAATCAGCTCC 226
DB 142644 GTTCCAGATGCGCATCTTGCTCTATTTCTTCACATGTACAGCCACAATCAGCTCC 142703

QY 227 CTGTCTATGAGGACAGCCGAGCGCATGACAGTCTGCTGTTCTGCTTCCGCGCTCT 286
DB 142704 CTGTCTATGAGGACAGCCGAGCGCATGACAGTCTGCTGTTCTGCTTCCGCGCTCT 142763

QY 287 GGCTCTTCTTTTGTGGGAGGCTTTTGGGACAGCAGATGTTTCTACCTGATGAGG 346

```
Db 142764 GGGCTCTCTTTGGGGGAGCTTTTGGGAGACACCAAGATGTTTCTACGTATGAG 142823
QY 347 CCGCCCTGGACGCTGTGGATGACCTTCAACAGAGACCCGCTGTGATGCTGACATGCT 406
Db 142824 CCGCCCTGGACGCTGTGGATGACCTTCAACAGAGACCCGCTGTGATGCTGACATGCT 142883
QY 407 GTGGGAGATGATACGGGCGCTTCTGTGACATGAGACGCTTGTGATGCTGATGCT 466
Db 142884 GTGGGAGATGATACGGGCGCTTCTGTGACATGAGACGCTTGTGATGCTGATGCT 142943
QY 467 GAACCTGGTCCCGGAGACAGTCCAGCCTTTTCAAGTGGAGAAACACCCGGGCGCTGTG 526
Db 142944 GAACCTGGTCCCGGAGACAGTCCAGCCTTTTCAAGTGGAGAAACACCCGGGCGCTGTG 143003
QY 527 TCGACACCTGCTGTGATCAATCCCAAGATGAAATCAATCCCGGCGCTGCTACAGCAG 586
Db 143004 TCGACACCTGCTGTGATCAATCCCAAGATGAAATCAATCCCGGCGCTGCTACAGCAG 143063
QY 587 CTCTGTGACAGTCAACAGCCTTGTGAGTGGAGAGAGCGCTCCGCTCTACAGCAGC 646
Db 143064 CTCTGTGACAGTCAACAGCCTTGTGAGTGGAGAGAGCGCTCCGCTCTACAGCAGC 143123
QY 647 GTGGTGTCAAGAGAGTGGCGCTTCTTCACTGACAGTCCCTTCAACCCGCTGTGAAAGAC 706
Db 143124 GTGGTGTCAAGAGAGTGGCGCTTCTTCACTGACAGTCCCTTCAACCCGCTGTGAAAGAC 143183
QY 707 CCGTCCCTCAACCTGCAATCTGTGACAGTGGTCCGGGAGCCCGGCGCTGTGCTTCC 766
Db 143184 CCGTCCCTCAACCTGCAATCTGTGACAGTGGTCCGGGAGCCCGGCGCTGTGCTTCC 143243
QY 767 CGAAGACGACAAAGAGAGATCTCATGATTGACAGTGGCATTTGATGGGGACAGATGAG 826
Db 143244 CGAAGACGACAAAGAGAGATCTCATGATTGACAGTGGCATTTGATGGGGACAGATGAG 143303
QY 827 CAAAACTCAAGAGAGAGACCAACCTACTATGTGATGAGATCTGTGCAAAAGCAG 886
Db 143304 CAAAACTCAAGAGAGAGACCAACCTACTATGTGATGAGATCTGTGCAAAAGCAG 143363
QY 887 CTGAGATCTTCAAGAGACATCTGCTTGTGCAAGGCGCTGTGCAAGAGCCTTACTGCTT 946
Db 143364 CTGAGATCTTCAAGAGACATCTGCTTGTGCAAGGCGCTGTGCAAGAGCCTTACTGCTT 143423
QY 947 GTGGCTATGAGACCTGTGCTGAGAGCCCTGTGGCCAGAGACTTCCGAATGTATGAATTC 1006
Db 143424 GTGGCTATGAGACCTGTGCTGAGAGCCCTGTGGCCAGAGACTTCCGAATGTATGAATTC 143483
QY 1007 GTGGGATTTGAAATCTTGGCCCATCTTCAAGACCTGGGTGATACATCACCCGAGCAG 1066
Db 143484 GTGGGATTTGAAATCTTGGCCCATCTTCAAGACCTGGGTGATACATCACCCGAGCAG 143543
QY 1067 GGCATGGGTGATCCAGCCTTTCACACAAATGCGAGGAGATCCCTTAATGTCTCCAGCCT 1126
Db 143544 GGCATGGGTGATCCAGCCTTTCACACAAATGCGAGGAGATCCCTTAATGTCTCCAGCCT 143603
QY 1127 TGGGCGTGTGCTTGGCCCTTGAAGAGTTTCTGACCTTGAAGAGCCTGTGGGATGCC 1186
Db 143604 TGGGCGTGTGCTTGGCCCTTGAAGAGTTTCTGACCTTGAAGAGCCTGTGGGATGCC 143663
QY 1187 ATGAATTTGCTGGGCTACCGCAGCTGATCTGAGAACAGAAACAGAAACCTTGTCTG 1246
Db 143664 ATGAATTTGCTGGGCTACCGCAGCTGATCTGAGAACAGAAACAGAAACCTTGTCTG 143723
QY 1247 GATCTTGTCTTACCTGAGATCTGCTTCCAGCAATTCACATTAAGAGGCTTGAAGGCTTT 1306
Db 143724 GATCTTGTCTTACCTGAGATCTGCTTCCAGCAATTCACATTAAGAGGCTTGAAGGCTTT 143783
QY 1307 GCTGCCACCTGTGTGACCTTCACTTCTGTGATGCTTGTGAGCCTTGTGCTTCAAT 1366
Db 143784 GCTGCCACCTGTGTGACCTTCACTTCTGTGATGCTTGTGAGCCTTGTGCTTCAAT 143843
QY 1367 CTCTGAGCCCTTAACTATCTGTGTGGGTATCACACTGAGTGTGTGTCCACAGCT 1426
```

```
Db 143844 CTCTGAGCCCTTAACTATCTGTGTGGGTATCACACTGAGTGTGTGTCTCCACAGCT 143903
QY 1427 GCTCAAGCAGAGAGACTTGTGTGTCATGCTTGTGTAGAAAAACAGCTGGGAACTT 1486
Db 143904 GCTCAAGCAGAGAGACTTGTGTGTCATGCTTGTGTAGAAAAACAGCTGGGAACTT 143963
QY 1487 ATGTGACAGACATCTCCACAGTGAAGAGGATGTCTTCTTCTTCTTGTGATCTT 1546
Db 143964 ATGTGACAGACATCTCCACAGTGAAGAGGATGTGTCTTCTTCTTGTGATCTT 144023
QY 1547 CCGTGTGGGAGACTTCCAGAGACTTGTGGCTGTGAGGCTTATTAAGACAGACAGTA 1606
Db 144024 CCGTGTGGGAGACTTCCAGAGACTTGTGGCTGTGAGGCTTATTAAGACAGACAGTA 144083
QY 1607 TCAAGTGAATTTGATCAATTAACCTCCCTGTCCATATCTTGGCAATGGGAAATGATCTT 1666
Db 144084 TCAAGTGAATTTGATCAATTAACCTCCCTGTCCATATCTTGGCAATGGGAAATGATCTT 144143
QY 1667 TCACCAAGAGAGCTCCAGAGATTTTCCACAGAGATGCGAAATTCGAGCCCTTGGAGTTC 1726
Db 144144 TCACCAAGAGAGCTCCAGAGATTTTCCACAGAGATGCGAAATTCGAGCCCTTGGAGTTC 144203
QY 1727 CAATGGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1786
Db 144204 CAGT-GGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 144262
QY 1787 T-ACAGTATCGGTAATCAGAAATATGAACAAATATCTG-ACAAAGAGAGAGAGAG 1844
Db 144263 TCACAGTATCGGTAATCAGAAATATGAACAAATATCTG-ACAAAGAGAGAGAGAGAG 144322
QY 1845 AAGTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1903
Db 144323 AAGTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 144382
QY 1904 CATGAAACTTTGACCTGTGAACCTTCCATCTGTATTAATTAATTAATTAATTAATTA 1963
Db 144383 CATGAAACTTTGACCTGTGAACCTTCCATCTGTATTAATTAATTAATTAATTAATTA 144441
QY 1964 TTTGTTTAAAGATCTCCCTTTTATGCTTCTTATTTAGCAATTAATTAATTAATTA 2023
Db 144442 TTTGTTTAAAGATCTCCCTTTTATGCTTCTTATTTAGCAATTAATTAATTAATTA 144501
QY 2024 TGGGATCTTAAAAAATAA 2042
Db 144502 TGGGATCTTAAAAAATAA 144520
```

RESULT 7
AC138848/c
AC138848 206943 bp DNA linear HTG 21-JAN-2003
LOCUS
DEFINITION Homo sapiens chromosome 5 clone RP11-1301B21, WORKING DRAFT
SEQUENCE, 7 unordered pieces.
AC138848
AC138848.1 GI:27805260
VERSION
KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT; HTGS-ACTIVEFIN.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 206943)
DOE Joint Genome Institute.
TITLE Direct Submissions
JOURNAL Submitted (21-JAN-2003) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 2719263
Center clone name: RPCI-11.1301B21

Summary Statistics

Consensus quality: 202039 bases at least Q40
Consensus quality: 202804 bases at least Q30
Consensus quality: 203337 bases at least Q20
Estimated insert size: 175000; agarose-fp estimation
Estimated insert size: 206343; sum-of-coverage estimation
Quality coverage: 15.28 in Q20 bases; agarose-fp estimation
Quality coverage: 12.96 in Q20 bases; sum-of-coverage estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1219: contig of 1219 bp in length
* 1220 1319: gap of unknown length
* 1320 2493: contig of 1174 bp in length
* 2494 2593: gap of unknown length
* 2594 3667: contig of 1073 bp in length
* 3667 5167: gap of unknown length
* 5167 5267: gap of 1401 bp in length
* 5267 12818: gap of unknown length
* 12818 12919: contig of 7551 bp in length
* 12919 36169: gap of unknown length
* 36169 36270: contig of 23251 bp in length
* 36270 206943: gap of unknown length
Location/Qualifiers
1. 206943

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RPI1-1301B21"
/clone_lib="RPCI human BAC library 11"

BASE COUNT 56312 a 44122 c 46163 g 59225 t 1121 others

Query Match 90.9%; Score 1856.6; DB 2; Length 206943;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1920; Conservative 0; Mismatches 14; Indels 5; Gaps 5;

107 AAGGCTTCACCTTACGACAAATGCTACTGCTTAAATGAAGCTCGTGTTCG 166
125368 AAGGCTTCACCTTACGACAAATGCTACTGCTTAAATGAAGCTCGTGTTCG 125309
167 GTTTCAGATGAGGATCTTGGCTCTATCTTCCATGTATACAGCCAAACATCAGCTCC 226
125308 GTTTCAGATGAGGATCTTGGCTCTATCTTCCATGTATACAGCCAAACATCAGCTCC 125249
227 CTTGCTATTAAGGACAGCCGAGGACATGACAGTCTGCTTCTCTGGGCTCT 286
125248 CTTGCTATTAAGGACAGCCGAGGACATGACAGTCTGCTTCTCTGGGCTCT 125189
287 GGCTCTCTTTTGGGGAGGCTTTTGGGGAGGACCCAGATGTTTCTACCGTATGAG 346
125188 GGCTCTCTTTTGGGGAGGCTTTTGGGGAGGACCCAGATGTTTCTACCGTATGAG 125129
347 CCCGCTTGACAGTGTGATGACCTTCAAGCAGAGACCGCTGATGCTGCATGCT 406
125128 CCCGCTTGACAGTGTGATGACCTTCAAGCAGAGACCGCTGATGCTGCATGCT 125069
407 GTGCGGATCTGATAGGAGGCTCTTCTTGTGCGACATGAGGCTTTTATGCTATG 466
125068 GTGCGGATCTGATAGGAGGCTCTTCTTGTGCGACATGAGGCTTTTATGCTATG 125009
467 GAACCTGCTCCCGAGAGACAGTCCAGCTCTTCACTGGGAGAAACGCCGCCCTGTGT 526

125008 GAACCTGCTCCCGAGAGACAGTCCAGCTCTTCACTGGGAGAAACGCCGCCCTGTGT 124949
527 TCGGACAGTCCCGTGTGATCATCCACAAAGATGAATATCCCGGGCTCATGTCAG 586
124948 TCGGACAGTCCCGTGTGATCATCCACAAAGATGAATATCCCGGGCTCATGTCAG 124889
587 CTCCTGTGACATCAACAGCCCTTGTGAGGTGAGAGAGCCCTGCTCTCAACAGCCAC 646
124888 CTCCTGTGACATCAACAGCCCTTGTGAGGTGAGAGAGCCCTGCTCTCAACAGCCAC 124829
647 GTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACATCCCTTACCCGCTGCTGAAGAC 706
124828 GTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACATCCCTTACCCGCTGCTGAAGAC 124769
707 CCCTCCCTCAACCTGATATCGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 766
124768 CCCTCCCTCAACCTGATATCGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 124709
767 CGAGAACGACAAAGGAGATCTCATGATGACATGCTGCAATGATGAGGAGCAGATGAG 826
124708 CGAGAACGACAAAGGAGATCTCATGATGACATGCTGCAATGATGAGGAGCAGATGAG 124649
827 CAAAACCTGAAGAGAGAGACCAACCTCTATGATGATGATGATGATGATGATGATGATGAT 886
124648 CAAAACCTGAAGAGAGAGACCAACCTCTATGATGATGATGATGATGATGATGATGATGAT 124589
887 CTGAGATCTACAAAGACATCCAGTCTTGGCCAAAGCCCTGAGAGAGCTACTGCTT 946
124588 CTGAGATCTACAAAGACATCCAGTCTTGGCCAAAGCCCTGAGAGAGCTACTGCTT 124529
947 GTGCGCTATGAGAGCTGCTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1006
124528 GTGCGCTATGAGAGCTGCTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 124469
1007 GTGCGATGGAATTTCTTGGCCCATCTTCAAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1066
124468 GTGCGATGGAATTTCTTGGCCCATCTTCAAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 124409
1067 GGCATGGGTGACACGCTTTCACACAAATGCGAGGATGCCCTTAATGCTGCCAGCT 1126
124408 GGCATGGGTGACACGCTTTCACACAAATGCGAGGATGCCCTTAATGCTGCCAGCT 124349
1127 TGGCGCTGCTTGGCCCATCTTCAAGAGCTTCTGCACTTCAAGAAACCTGCTGCTGCT 1186
124348 TGGCGCTGCTTGGCCCATCTTCAAGAGCTTCTGCACTTCAAGAAACCTGCTGCTGCT 124289
1187 ATGAATTTGCTGAGGCTGACGCGCAGCTGATGATGATGATGATGATGATGATGATGATGAT 1246
124288 ATGAATTTGCTGAGGCTGACGCGCAGCTGATGATGATGATGATGATGATGATGATGATGAT 124229
1247 GATCTTCTGCTACCTGACCTGCTCCTGAGCAATTCACATAGAGGCTTGAAGAGCTTT 1306
124287 GATCTTCTGCTACCTGACCTGCTCCTGAGCAATTCACATAGAGGCTTGAAGAGCTTT 124169
124288 GATCTTCTGCTACCTGACCTGCTCCTGAGCAATTCACATAGAGGCTTGAAGAGCTTT 124169
1307 GCTGCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1366
124168 GCTGCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 124109
1367 CTTGAGGCTTAACTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1426
124108 CTTGAGGCTTAACTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 124049
1427 GCTCAAGCAGAGGATTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1486
124048 GCTCAAGCAGAGGATTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 123989
1487 ATGTGAGCAGCAATCCACCAAGTGAAGAGGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1546
123988 ATGTGAGCAGCAATCCACCAAGTGAAGAGGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 123929
1547 CCTGTCTGGGAGCTTCAAGACATTTTGGGCTGAGGCTTATTAAGACAGACACAGTA 1606
CCTGTCTGGGAGCTTCAAGACATTTTGGGCTGAGGCTTATTAAGACAGACACAGTA

THIS PAGE BLANK (USPTO)

Db 1276 CATACTAACGCCAGAAAGCCCTCAGCTCTCAGGGCTGGCTGCTTACTTAC 1335
OY 1148 GAAAGGTTCTCAGATTCAAAAGCCGTGGCCATGCCATGAATTTCTGGCTACCGC 1207
Db 1336 GAAAGGTTCTCAGATTCAAAAGCCGTGGCTGAGCTATGATTTGCTGGATACCTC 1395
OY 1208 CACGTCAATCTGAACAAGAACAGAAACCTGTTGCTGATCTTCTGTACTGAGCT 1267
Db 1396 CAGGTCAATCTGAACAAGAACAGAAACCTGTTGCTGATCTTCTGTACTGAGCT 1455
OY 1268 GTCCCTGAGCAATCCACTAAGAGGTTGAGAAGGCTTGTCTG--CCACTGGTGTACG 1325
Db 1456 ATCTTGGGGCAGGCTTCCGAGAGGTTAAGAGGCTGTCTGCACCCCTTGGTCCAGC 1515
OY 1326 CTCAGTCACTTCTCTGAATGCTTCTGAGCCTTGCCCTACATCTGAGCCTTAATACAT 1385
Db 1516 CTTAGTCACT--CATTAACGACGAGAGCCTTAAGTAGAACCAACTGAGTGCCCTT 1572
OY 1386 GTCTGTGGGTATCACACTGAGTGTGATTTGTCCACAGTGTCAAGCAGAGGACTTT 1445
Db 1573 TCTCCTAGGCCCAAGCAGAGAGGCTTGTGTCTATCTCAGGCTTACCTACACTGA 1632
OY 1446 TGTGTCCATGCTTGTGTAGAAAACAGACTGGGGAACCTTAATGTAGACAGACATCCA 1505
Db 1633 GCTTAAAAAGCCAAAGAAACATATCTTCTGTGAAATACTTAAGAACCTTAAGCAG 1692
OY 1506 CCAGTGAACAGGGTA--TTGCTCTTCTTCTTCTTCTGATCTTCTGCTGGGAGACT 1562
Db 1693 CCCCTTGAACCTGTCAAGCAAGACTTCTTGAATTTGGCTTCTTACTGTGATAC 1752
OY 1563 TCAGAGACTTGTGGCTGAGGCTTATTAAAGCAGACAGATATCACTGATGATTTGATCC 1622
Db 1753 TTGAGACTAGGCTGTGAGGCTTACTGACACAGCAAGCAATCCATGAGTGTCT 1812
OY 1623 ATTAACCTCCCTGTCCATATCTTGGCCAAATGGGAATGATCTTCAACCAAGAGCTCAC 1682
Db 1813 GTAAACCTCCCTGTCCATATCTTCCGAAAGAAAGGGAGAAATAAAGCCCAAGGAAATG 1872
OY 1683 CAGCATTTTCCACAGATGCGA-----ATTCTGAGCCCTTGGAGTTCCCA 1729
Db 1873 CGGCTCTTACCAAGAGCTCCCAAGAACTTCCACAGGACATAGACCTCAAGTTCCA 1932
OY 1730 TGGATTCAGGAAGAGAGTGGAAACAAGTTGATGCTTACTATGAGCTTGACCATAC 1789
Db 1933 GCGGATTTGAGAGAGCAGAGAGCGGGGTGGCTTATCT--GTGAATCTGTGCCATGT 1991
OY 1790 AGCTATGCTAATCAGAAATATGAACAAATCTGTGACAAAGAGCAAGCTTAAGTT 1849
Db 1992 CACTATCCGTATATCAGAAACAGAGCAAA--CAACACAAACAGGGGCAAGACTCAAGTT 2048
OY 1850 CACAAGTGCTGGGCTTGAT 1870
Db 2049 CAAGGCTGTCTGGATGAT 2069

Search completed: August 14, 2003, 15:34:38
Job time : 7298.74 secs

```

|||||
Db 455 CCGGGCCCTGTTCTGACACCTGCTGATCATCAACCAAGAAATATATCCCG 514
OY 574 GGCTACATCAGAGCTCTCTGTGAGTCAACAGCCCTTTGAGTGTGAGAGAGCCGCC 633
Db 515 GGCTACATCAGAGCTCTCTGTGAGTCAACAGCCCTTTGAGTGTGAGAGAGCCGCC 574
OY 634 CTCCTACAGCAGCTGTGTCTCAAGAGAGTGGCTTTTCAACCTGACATCCCTTACCC 693
Db 575 CTCCTACAGCAGCTGTGTCTCAAGAGAGTGGCTTTTCAACCTGACATCCCTTACCC 634
OY 694 GTGCGAAGAGACCTCTCCCTCACTGATATCGGACCTGTGTGAGAGAGAGAGAGAGAG 753
Db 635 GTGCGAAGAGACCTCTCCCTCACTGATATCGGACCTGTGTGAGAGAGAGAGAGAGAG 694
OY 754 CGTGTTCCTGTCCTCCGAGAGACGACAAAGAGAGATCATGATGACATGAGTGTGAT 813
Db 695 CGTGTTCCTGTCCTCCGAGAGACGACAAAGAGAGATCATGATGACATGAGTGTGAT 754
OY 814 GGGGAGCATGTGACAAACAACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 872
Db 755 GGGGAGCATGTGACAAACAACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 814
OY 873 -TCGTGCAAAAGCAGC-TGAGATCTACAAAGAC 904
Db 815 TCTGGCCAAAGCAGCTTGAGAGACTACAAAGAC 848

RESULT 15
AF109155 2201 bp mRNA linear ROD 27-JUL-1999
LOCUS Mus musculus L-selectin ligand sulfotransferase mRNA, complete cds.
DEFINITION AF109155
ACCESSION AF109155.1 GI:5596405
VERSION
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
1 (bases 1 to 2201)
Hirooka,N., Petryniak,B., Nakayama,J., Teubol,S., Suzuki,M.,
Yeh,J.C., Izawa,D., Tanaka,T., Miyasaka,M., Lowe,J.B. and Fukuda,M.
a novel, high endothelial venule-specific sulfotransferase
expresses 6-sulfo sialyl Lewis(x), an L-selectin ligand displayed
by CD34
Immunology 11 (1), 79-89 (1999)
JOURNAL
MEDLINE 99361934
PUBMED 10435581
REFERENCE 2 (bases 1 to 2201)
AUTHORS Hirooka,N. and Fukuda,M.
TITLE Direct Submission
JOURNAL Submitted (23-NOV-1998) Glycobiology, The Burnham Institute, 10901
North Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES
source
1..2201
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
319..1485
/function="directs expression of 6-sulfo sialyl Lewis x,
an L-selectin ligand displayed by CD34"
/note="LST"
/codon_start=1
/protein_id="A045579.1"
/db_xref="GI:5596406"
/translation="MMILKKGRLLMFISQVIVVAFITMVSVHRLSORESRPVH
LVLSWRSGSSFTVGLFGCHPVFLMEPAHVMTTSTIMKLMHARLDLRVFL
CDMSVFDAYMNPGRKSSLFQWESRALCSAPVCDFFPAHISSPKCKILCGQPF
DMVEKACRSHGVNVEKVFSLALYPLDPSLNHVHVRPRAVFRSREHTTI
ELVVDSHIVLQGLETIKEEDPYAMKIKSQSDIYVAIQTLPEALQOQRYLFLYE
DLVRAPIAOTRLYKFGVGLDFPHLOTGVHNTVRGKMGQOHAFNLNARALNVSQAMR
WSLPEKVSQLODAGSAMDLGLTGLQVRSQBOGNLSLDLSSSHLQGVREG"

```

```

BASE COUNT 537 a 629 c 534 g 501 t
ORIGIN
Query Match 37.0%; Score 755; DB 10; Length 2201;
Best Local Similarity 67.2%; Pred. No. 8.7e-180;
Matches 1183; Conservative 0; Mismatches 550; Indels 28; Gaps 7;
OY 128 ATGCTACTGCTTAAATAAAGAGCTCTGCTGTTCTGTTTCCAGATGACATCTTG 187
Db 319 ATGATGCTGTTGAAGAAAGAGAGGCTGTGATGTTCTGTTTCCAGATGACATCTTG 378
OY 188 GCTCATTTTTCACATGTACAGCCCAACATACAGTCCCTGTCTATGAAGACAGCC 247
Db 379 GCTCTCTTATCATATGTCCTCCACACACACTTTCCTCAAGAGAGAGATGCAGAG 437
OY 248 GAGCGCATGAGCTGTGTTCTGTTCTGTTCTGAGGCTGTGAGCTTCTTTTGGGAG 307
Db 438 -GCCGCTCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 495
OY 308 CTTTTCGGGACACACCATGATGTTTCTACATGATGAGCCCGCTGACATGTGATG 367
Db 496 CTTTTCGGGACACACCATGATGTTTCTACATGATGAGCCCGCTGACATGTGATG 555
OY 368 ACCTTCACAGACACACCATGATGTTTCTACATGATGAGCCCGCTGACATGTGATG 427
Db 556 ACCTTCACAGACACACCATGATGTTTCTACATGATGAGCCCGCTGACATGTGATG 615
OY 428 GTCTTCCTGTGACATGAGCGCTTGTATGCTTACATGAACTGTCCTCCGAGACAG 487
Db 616 GTCTTCCTGTGACATGAGCGCTTGTATGCTTACATGAACTGTCCTCCGAGACAG 675
OY 488 TCCAGCTCTTTCAGTGGAGAACACCGGCGCTGTGTTCTGACATGCTGTGACATG 547
Db 676 TCCAGCTCTTTCAGTGGAGAACACCGGCGCTGTGTTCTGACATGCTGTGACATG 735
OY 548 ATCCACAGATGAATATCCCGGCGCTGTGTTCTGACATGCTGTGACATGCTGTGAC 607
Db 736 TCCCGTCCGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 795
OY 608 TTTGAGTGTGAGAGAGCGCTGCGCTCTACAGCAGAGTGTGTCAAGAGAGTGGC 667
Db 796 TTTGAGTGTGAGAGAGCGCTGCGCTCTACAGCAGAGTGTGTGTGTGTGTGTGTGT 855
OY 668 TTTCTACCTGAGTGTGCTTACCGCGCTGTGAAAGACCCCTGCTCAACCTGATATC 727
Db 856 TTTCTACCTGAGTGTGCTTACCGCGCTGTGAAAGACCCCTGCTCAACCTGATATC 915
OY 728 GTGACACTGTGCGGAGACCCCGGCGCTGTTCCCGAGAACGCAAAAGGAGAT 787
Db 916 GTGACACTGTGCGGAGACCCCGGCGCTGTTCCCGAGAACGCAAAAGGAGAT 975
OY 788 CTCATGATTTGACATGCTATGTTGATGGGAGCATGACCAAAACTCAAGAGAGAG 847
Db 976 CTCGATGTTGACATGCTATGTTGATGGGAGCATTTGGAAACATCAAGAGAGAG 1035
OY 848 CAACCTACTATGTGATGAGTCAATCTGCAAAAGCAGCTGAGATCTTCAAGACATC 907
Db 1036 CAGCCCTATTATGCCATGAAGATCATCTCAAAAGCAGCTGAGATCTTCAAGACATC 1095
OY 908 CAGTCTTCCCAAGGCGCTGAGAAAGCTACCTGCTGTGCGTATAGAACCTGGGT 967
Db 1096 CAACCTCTCTGAGAGCTGTGAGAGCGCTACCTGCTGTGCGTATAGAACCTGGGT 1155
OY 968 CGAGCCCTGTGCGGACAGCTTCCGAAAGTATGATGTTGTTGATGATGATGATGATG 1027
Db 1156 CGGAGACCCCTGCGGACAGACAGACATATTAATTTGTGGGTTGATTTTGGCC 1215
OY 1028 CATCTTACAGCTGGGTGATTAATCAATCAACCGAGAGAGAGATGGGTACACGCTTC 1087
Db 1216 CAGCTTCACAAACATGTTTACATGATGATGATGATGATGATGATGATGATGATG 1275
OY 1088 CACACAAAGGAGAGATGCTTATATGCTTCCAGGCTGTGCGGCTGTGCGCTT 1147
Db 1147 CACACAAAGGAGAGATGCTTATATGCTTCCAGGCTGTGCGGCTGTGCGCTT 1147

```

FEATURES FT Location/Qualifiers
 source 1..877
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

BASE COUNT 183 a 266 c 221 g 203 t 4 others
 ORIGIN

Query Match 37.4% Score 763.6; DB 6; Length 877;
 Best Local Similarity 97.7%; Pred. No. 5.3e-182;
 Matches 795; Conservative 0; Mismatches 16; Indels 3; Gaps 2;

94 AAGCCAAAGCAGAGTCTTCCACTTCAGCACAATGCTACTGCTTAAATAATGAACT 153
 |||||
 35 AAGCCGCTCTGCAAGGCTTCCACTTCAGCACAATGCTACTGCTTAAATAATGAACT 94
 |||||
 154 CCTGCTGTTCTGGTTTCCAGATGCGCATCTGGCTCTATTCTTCCACATGTACGCCA 213
 |||||
 95 CCTGCTGTTCTGGTTTCCAGATGCGCATCTGGCTCTATTCTTCCACATGTACGCCA 154
 |||||
 214 CAACATCAGTCCCTGCTATGAGGAGCAGCCGAGGCATGCACGCTGCTGCTGCTC 273
 |||||
 155 CAACATCAGTCCCTGCTATGAGGAGCAGCCGAGGCATGCACGCTGCTGCTGCTC 214
 |||||
 274 TTCTGGGCGCTGGCTCTTCTTTTGGGGCAGCTTTTGGGAGCAGCCAGATGTTT 333
 |||||
 215 TTCTGGGCGCTGGCTCTTCTTTTGGGGCAGCTTTTGGGAGCAGCCAGATGTTT 274
 |||||
 334 CTACCTGATGAGCCCGCTGGCAGCTGTGATGATCCTTCAAGCAGACCCGCTGAT 393
 |||||
 275 CTACCTGATGAGCCCGCTGGCAGCTGTGATGATCCTTCAAGCAGACCCGCTGAT 334
 |||||
 394 GCTGACATGCTGCTGGGAGTGTATGATGAGGCGCGCTCTTCTTGGCAGATGAGGCT 453
 |||||
 335 GCTGACATGCTGCTGGGAGTGTATGATGAGGCGCGCTCTTCTTGGCAGATGAGGCT 394
 |||||
 454 TGATGCTACATGGAACCTGTGCTCCCGAGAGACAGTCCAGCCCTTTCAGTGGAGAACAG 513
 |||||
 395 TGATGCTACATGGAACCTGTGCTCCCGAGAGACAGTCCAGCCCTTTCAGTGGAGAACAG 454
 |||||
 514 CCGGGCCCTGTGTTCTGACCTGCTGTGACATATCCCAAGATGAATCATTCCCG 573
 |||||
 455 CCGGGCCCTGTGTTCTGACCTGCTGTGACATATCCCAAGATGAATCATTCCCG 514
 |||||
 574 GGTCTACTGAGGCTCTGCTGACAGTCAACAGCCCTTGTAGTGGTGGAGAGCCCTGCG 633
 |||||
 515 GGTCTACTGAGGCTCTGCTGACAGTCAACAGCCCTTGTAGTGGTGGAGAGCCCTGCG 574
 |||||
 634 CTCTACAGCAGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTTACCC 693
 |||||
 575 CTCTACAGCAGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTTACCC 634
 |||||
 694 GCTCTGAAGAGCCCTCCCTCAACCTGATATGCTGACCTGCTGCGGAGCCCGGCG 753
 |||||
 635 GCTCTGAAGAGCCCTCCCTCAACCTGATATGCTGACCTGCTGCGGAGCCCGGCG 694
 |||||
 754 CGTGTTCGTTCCCGAGAGCCAGAAAGGAGATCTCATGTGATGACATTCGATTTGNAT 754
 |||||
 695 CGTGTTCGTTCCCGAGAGCCAGAAAGGAGATCTCATGTGATGACATTCGATTTGNAT 754
 |||||
 814 GGGGAGCATGAGCAAAAATCTCAAGAGAGAGACCAACCTACTATGTGATGACAGTCA - 872
 |||||
 755 GGGGAGCATGAGCAAAAATCTCAAGAGAGAGACCAACCTACTATGTGATGACAGTCA 814
 |||||
 873 -TTCGCAAGACAGC-TGAGATCTACAGACC 904
 |||||
 815 TCTGGCCAAAGCCAGCTTGGAGAACTACAGAAC 848
 |||||

RESULT 14
 BDI26950 877 bp DNA linear PAT 18-SEP-2002

DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BDI26950
 VERSION BDI26950.1 GI:23221895
 KEYWORDS JP 2002017375-A/2381.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens (human)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

OS Homo sapiens (human)
 PN JP 2002017375-A/2381
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
 SHINICHI KOUJIMA,
 PI TETSUJI OTSUKI,HISASHI KOGA
 PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/00
 PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key

FEATURES FT Location/Qualifiers
 source 1..877
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

BASE COUNT 183 a 266 c 221 g 203 t 4 others
 ORIGIN

Query Match 37.4% Score 763.6; DB 6; Length 877;
 Best Local Similarity 97.7%; Pred. No. 5.3e-182;
 Matches 795; Conservative 0; Mismatches 16; Indels 3; Gaps 2;

94 AAGCCAAAGCAGAGTCTTCCACTTCAGCACAATGCTACTGCTTAAATAATGAACT 153
 |||||
 35 AAGCCGCTCTGCAAGGCTTCCACTTCAGCACAATGCTACTGCTTAAATAATGAACT 94
 |||||
 154 CCTGCTGTTCTGGTTTCCAGATGCGCATCTGGCTCTATTCTTCCACATGTACGCCA 213
 |||||
 95 CCTGCTGTTCTGGTTTCCAGATGCGCATCTGGCTCTATTCTTCCACATGTACGCCA 154
 |||||
 214 CAACATCAGTCCCTGCTATGAGGAGCAGCCGAGGCATGCACGCTGCTGCTGCTC 273
 |||||
 155 CAACATCAGTCCCTGCTATGAGGAGCAGCCGAGGCATGCACGCTGCTGCTGCTC 214
 |||||
 274 TTCTGGGCGCTGGCTCTTCTTTTGGGGCAGCTTTTGGGAGCAGCCAGATGTTT 333
 |||||
 215 TTCTGGGCGCTGGCTCTTCTTTTGGGGCAGCTTTTGGGAGCAGCCAGATGTTT 274
 |||||
 334 CTACCTGATGAGCCCGCTGGCAGCTGTGATGATCCTTCAAGCAGACCCGCTGAT 393
 |||||
 275 CTACCTGATGAGCCCGCTGGCAGCTGTGATGATCCTTCAAGCAGACCCGCTGAT 334
 |||||
 394 GCTGACATGCTGCTGGGAGTGTATGATGAGGCGCGCTCTTCTTGGCAGATGAGGCT 453
 |||||
 335 GCTGACATGCTGCTGGGAGTGTATGATGAGGCGCGCTCTTCTTGGCAGATGAGGCT 394
 |||||
 454 TGATGCTACATGGAACCTGTGCTCCCGAGAGACAGTCCAGCCCTTTCAGTGGAGAACAG 513
 |||||
 395 TGATGCTACATGGAACCTGTGCTCCCGAGAGACAGTCCAGCCCTTTCAGTGGAGAACAG 454
 |||||
 514 CCGGGCCCTGTGTTCTGACCTGCTGTGACATATCCCAAGATGAATCATTCCCG 573
 |||||

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 93..1253
 /note="L5S1"
 /codon_start=1
 /product="L-selectin ligand sulfotransferase"
 /protein_id="AAK48417.1"
 /db_xref="GI:13897504"
 /translation="MLPKMKILFLIVSOMAILALFFHMSHNISLSMKRQPERM
 VLIVSRSRSGSFVQLFGQHDVFLYLPANHWMTKOSTAMLMHAWDLIRAV
 LCDMSVDAIMEPPGRQSLFQWENSRLSAPACDIIPODEIIIPRAHCSQOP
 FEVEKACRSYSHVLAKEVREFNLQSLPLKDPISLNIHVLVDPRVAFSRRTK
 GDLMDISRIVMGHOHLKEDQPYVMQICOSOLEIKYITQSLPKALOERYLLVRY
 EDIARAPVQTSRMEFVGLFELPHLOTWVNIITRGKNGDIAFHTNADLVNSQAW
 RMSLPYEVSRLOKACGDAMNLIQYRHRSEBQENLILDLISTVTEYEQIH"

BASE COUNT 286 a 393 c 336 g 318 t
 ORIGIN

Query Match 61.8%; Score 1262; DB 9; Length 1333;
 Best Local Similarity 100.0%; Pred. No. 3.6e-308;
 Matches 1262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 AAGGCTTCCACTTATGACACATGCTACGCTTAAAAAATGAGCTCTGCTGTTCTG 166
 DB 72 AAGGCTTCCACTTATGACACATGCTACGCTTAAAAAATGAGCTCTGCTGTTCTG 131
 QY 167 GTTTCACAGATGGCATCTGGCTATCTTCACATGTACAGCCAAACATCAAGCTCC 226
 DB 132 GTTTCACAGATGGCATCTGGCTATCTTCACATGTACAGCCAAACATCAAGCTCC 191
 QY 227 CTGTCTATGAGCAGACAGCCGAGCGCATGCACTGCTGTTCTGCTTCTGCGCTCT 286
 DB 192 CTGTCTATGAGCAGACAGCCGAGCGCATGCACTGCTGTTCTGCTTCTGCGCTCT 251
 QY 287 GGCTCTCTTGTGGGGAGCTTTTGGGAGCAGCCAGATGTTTCTACTGATGAGAG 346
 DB 252 GGCTCTCTTGTGGGGAGCTTTTGGGAGCAGCCAGATGTTTCTACTGATGAGAG 311
 QY 347 CCCGCTGACAGCTGTGATGACCTTCAAGCAGACAGCCGCTGATGCTGACATGAGCT 406
 DB 312 CCCGCTGACAGCTGTGATGACCTTCAAGCAGACAGCCGCTGATGCTGACATGAGCT 371
 QY 407 GTGCGGATCTATACAGCGCGCTCTTCTGTGCGACATGAGCGCTTGTGACCTACATG 466
 DB 372 GTGCGGATCTATACAGCGCGCTCTTCTGTGCGACATGAGCGCTTGTGACCTACATG 431
 QY 467 GAACCTGGTCCCGGAGACATGCAAGCCCTTTCAGTGGAGAAACAGCCGGCCCTGTGT 526
 DB 432 GAACCTGGTCCCGGAGACATGCAAGCCCTTTCAGTGGAGAAACAGCCGGCCCTGTGT 491
 QY 527 TCTGACACTGCTGTGACATCCCAACAAGATGAATATCCCCGGGCTCACTGACAG 586
 DB 492 TCTGACACTGCTGTGACATCCCAACAAGATGAATATCCCCGGGCTCACTGACAG 551
 QY 587 CTCCTGTGACATCAACAGCCCTTGTGAGGTGTGAGAGAGCCCTGCTCTACAGCCAC 646
 DB 552 CTCCTGTGACATCAACAGCCCTTGTGAGGTGTGAGAGAGCCCTGCTCTACAGCCAC 611
 QY 647 GTGGGTCTCAAGGAGGTGGCTTCTTCAACCTGCACTGCCCTTACCCGCTGTGAAAGAC 706
 DB 612 GTGGGTCTCAAGGAGGTGGCTTCTTCAACCTGCACTGCCCTTACCCGCTGTGAAAGAC 671
 QY 707 CCCTCCCTCAACCTGATCTGTGACACTGCTCCGGAGACCCCGGGCCCTGTTCCGTTCC 766
 DB 672 CCCTCCCTCAACCTGATCTGTGACACTGCTCCGGAGACCCCGGGCCCTGTTCCGTTCC 731
 QY 767 CGAGAACGCAAAAGGAGATCTATGATGACAGTGTGATGATGGGAGAGCATGAG 826
 DB 732 CGAGAACGCAAAAGGAGATCTATGATGACAGTGTGATGATGGGAGAGCATGAG 791
 QY 827 CAAAACATCAAAAGGAGAGCAACCCCTACTATGTGATGACAGTCTATGCGCAAGCCAG 886
 DB 827 CAAAACATCAAAAGGAGAGCAACCCCTACTATGTGATGACAGTCTATGCGCAAGCCAG 886

DB 792 CAAAACATCAAAAGGAGAGCAACCCCTACTATGTGATGACAGTCTATGCGCAAGCCAG 851
 QY 887 CTGGAGATCTAAACACATCCAGTCTTGGCCCAAGCCCTCAGAGAACGTCACCTGCT 946
 DB 852 CTGGAGATCTAAACACATCCAGTCTTGGCCCAAGCCCTCAGAGAACGTCACCTGCT 911
 QY 947 GTGCGCTATGAGACCTGAGCTGAGACCCCTGTGGCCAGACACTTCCGAATGTATGAATTC 1006
 DB 912 GTGCGCTATGAGACCTGAGCTGAGACCCCTGTGGCCAGACACTTCCGAATGTATGAATTC 971
 QY 1007 GTGGGATTTGGAATTTCTTGGCCCATCTTTCAGACCTGGGTGCTCAATCAATCAGGAGCAAG 1066
 DB 972 GTGGGATTTGGAATTTCTTGGCCCATCTTTCAGACCTGGGTGCTCAATCAATCAGGAGCAAG 1031
 QY 1067 GGCATGAGGAGACACGCTTTCACACAAATGCGCAGGAGATGCCCTTAATGTCTCCAGGCT 1126
 DB 1032 GGCATGAGGAGACACGCTTTCACACAAATGCGCAGGAGATGCCCTTAATGTCTCCAGGCT 1091
 QY 1127 TGGCGCTGCTCTTGGCCCATATGAAAAGTTTCTGCACTTCAGAAAAGCCTGTGGCAGTGC 1186
 DB 1092 TGGCGCTGCTCTTGGCCCATATGAAAAGTTTCTGCACTTCAGAAAAGCCTGTGGCAGTGC 1151
 QY 1187 ATGAATTTGCTGGGCTACCGCCACGTCAGATCTGAACAAGAACAGAAACCTGTTGCTG 1246
 DB 1152 ATGAATTTGCTGGGCTACCGCCACGTCAGATCTGAACAAGAACAGAAACCTGTTGCTG 1211
 QY 1247 GATCTTCTGTACCTGAGACTGCTCCCTGAGCAATTCATAGAGAGGTTGAGAAGCTTT 1306
 DB 1212 GATCTTCTGTACCTGAGACTGCTCCCTGAGCAATTCATAGAGAGGTTGAGAAGCTTT 1271
 QY 1307 GCTGCCACCTGCTGTGACGCTCAGTCACTTCTGTGAATGCTTGTGAGCCTTGCTTACAT 1366
 DB 1272 GCTGCCACCTGCTGTGACGCTCAGTCACTTCTGTGAATGCTTGTGAGCCTTGCTTACAT 1331
 QY 1367 CT 1368
 DB 1332 CT 1333

RESULT 13
 BD124832
 LOCUS
 DEFINITION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 877)
 Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
 Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
 Koga,H.
 Primer for synthesizing full-length cDNA and use thereof
 Patent: JP 2002017375-A 263 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/263
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,MI RAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NGAI,PI
 SHINICHI KOJIMA,
 PI TETSUJI OTSUKI,HISASHI KOCA
 PC C12N15/09,C07R14/47,C07R16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
 10,
 PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 Location/Qualifiers
 FT source 1..877

Db 268 CAACATAGCTCCCTGCTATGAGGACAGCCGACGACATGACCTGCTGCTGTC 327
 Qy 274 TTCTGCGGCTGCGGCTCTTCTTGTGGGGACAGCTTTTGGGAGACCCAGATGTTT 333
 Db 328 TTCTGCGGCTGCTGCTCTTCTTGTGGGGACAGCTTTTGGGAGACCCAGATGTTT 387
 Qy 334 CTACCTGATGAGAGCCGCTGCGACAGCTGTGATGACCTTCAAGCAGAGACCCGCTGAT 393
 Db 388 CTACCTGATGAGAGCCGCTGCGACAGCTGTGATGACCTTCAAGCAGAGACCCGCTGAT 447
 Qy 394 GCTGCAATGCTGTGGGGATCTGATACGGGCGCTTCTTGTGCGACATGACCTT 453
 Db 448 GCTGCAATGCTGTGGGGATCTGATACGGGCGCTTCTTGTGCGACATGACCTT 507
 Qy 454 TGATGCTACATGGAACCTGTCGCCGAGACAGTCCAGCCTTTAGTGGGAGAACAG 513
 Db 508 TGATGCTACATGGAACCTGTCGCCGAGACAGTCCAGCCTTTAGTGGGAGAACAG 567
 Qy 514 CCGGGCCCTGTGTTCTGACCTGCTGACATATCCACAGATGAATCATCCCG 573
 Db 568 CCGGGCCCTGTGTTCTGACCTGCTGACATATCCACAGATGAATCATCCCG 627
 Qy 574 GGTCTACATGAGGCTCTGTGCACTCAACAGCCCTTTGAGTGGTGGAGAGCCCTGCG 633
 Db 628 GGTCTACATGAGGCTCTGTGCACTCAACAGCCCTTTGAGTGGTGGAGAGCCCTGCG 687
 Qy 634 CTCTACAGCAGCTGTGCTCAAGAGAGTGGCCTTTCAACCTGAGTCCCTCAACC 693
 Db 688 CTCTACAGCAGCTGTGCTCAAGAGAGTGGCCTTTCAACCTGAGTCCCTCAACC 747
 Qy 694 GCTGCTGAAAGACCCCTCCCTCAACCTGATATGTCACCTGTCGGGAGCCCGGCG 753
 Db 748 GCTGCTGAAAGACCCCTCCCTCAACCTGATATGTCACCTGTCGGGAGCCCGGCG 807
 Qy 754 CGTGTCCGTTCCCGAAGACGACAAAGGAGATCTATGATTTGACAGTGCATTTGAT 813
 Db 808 CGTGTCCGTTCCCGAAGACGACAAAGGAGATCTATGATTTGACAGTGCATTTGAT 867
 Qy 814 GGGGACATGAGAAAACCTCAAGAGAGAGACCAACCCACTATGATGAGAGTCAT 873
 Db 868 GGGGACATGAGAAAACCTCAAGAGAGAGACCAACCCACTATGATGAGAGTCAT 927
 Qy 874 CTGCGAAGCAGCTGAGATCTACAGACATCACTGCTGCGCAAGGCCCTGCAGGA 933
 Db 928 CTGCGAAGCAGCTGAGATCTACAGACATCACTGCTGCGCAAGGCCCTGCAGGA 987
 Qy 934 ACGCTACCTGCTGTGCTATGAGACCTGCTCGAGCCCTGTGCGCCAGACTTCCG 993
 Db 988 ACGCTACCTGCTGTGCTATGAGACCTGCTCGAGCCCTGTGCGCCAGACTTCCG 1047
 Qy 994 AATGATGAATGTGGGATTTGGAATCTTCCCATCTTCAAGCCCTGGGCTATACAT 1053
 Db 1048 AATGATGAATGTGGGATTTGGAATCTTCCCATCTTCAAGCCCTGGGCTATACAT 1107
 Qy 1054 CACCGAGGACAGGAGCATGAGTGCACAGCTTTCACACAAATCCAGGATGCCCTAA 1113
 Db 1108 CACCGAGGACAGGAGCATGAGTGCACAGCTTTCACACAAATCCAGGATGCCCTAA 1167
 Qy 1114 TGTCTCCAGGCTTGGCGTGTGCTTCCCTATGAAAGGTTTCTGACTTCAGAAAGC 1173
 Db 1168 TGTCTCCAGGCTTGGCGTGTGCTTCCCTATGAAAGGTTTCTGACTTCAGAAAGC 1227
 Qy 1174 CTGTGGGATGACATGAATTTGCTGGGCTACCGCAGCTAGATCTGAACAAAGACAG 1233
 Db 1228 CTGTGGGATGACATGAATTTGCTGGGCTACCGCAGCTAGATCTGAACAAAGACAG 1287
 Qy 1234 AAGCCTGTGCTGATCTTCTGTACTGAGCTGCTCCGAGCAATCCACTAAGAGGG 1293
 Db 1288 AAGCCTGTGCTGATCTTCTGTACTGAGCTGCTCCGAGCAATCCACTAAGAGGG 1347
 Qy 1294 TTGAGAGGCTTGTGCTGCACTGTGTCAGCCTCACTTCATCTTCTGATGCTTCTGA 1353

Db 1348 TTGAGAGGCTTTGCTGCCACCTGCTGTCAGCTCACTTCTCTGATGCTTCTGA 1407
 Qy 1354 GCTTGGCTACATCTGAGCCTTAACATGATGCTGGGATACACAGTGTAGT 1413
 Db 1408 GCTTGGCTACATCTGAGCCTTAACATGATGCTGGGATACACAGTGTAGT 1467
 Qy 1414 TGTGTCCACAGCTGCTCAAGCAGAGACTTTTGTGCATGCTTGTCTAGAAAACAG 1473
 Db 1468 TGTGTCCACAGCTGCTCAAGCAGAGACTTTTGTGCATGCTTGTCTAGAAAACAG 1527
 Qy 1474 ACTGGGACCTTATGTGAGAGACATCCACAGAGTGAAGAAAGGATTTGCTTCTC 1533
 Db 1528 ACTGGGACCTTATGTGAGAGACATCCACAGTGAAGAAAGGATTTGCTTCTC 1587
 Qy 1534 TTTTCTGATCTTCTGCTGCTGGGAGACTTCAAGACTTTTGTGCTGAGAGCCTATTAA 1593
 Db 1588 TTTTCTGATCTTCTGCTGCTGGGAGACTTCAAGACTTTTGTGCTGAGAGCCTATTAA 1647
 Qy 1594 GCACGACACAGTATCAGTGAATTCATTAACCTCCCTGTCACATCTTGGCCAAAG 1653
 Db 1648 GCACGACACAGTATCAGTGAATTCATTAACCTCCCTGTCACATCTTGGCCAAAG 1707
 Qy 1654 GGGATGATCTTTCACCAAGAGCTCACAGCATTTTCCACAGATGCGAATTCGAG 1713
 Db 1708 GGGATGATCTTTCACCAAGAGCTCACAGCATTTTCCACAGATGCGAATTCGAG 1767
 Qy 1714 CCTTGGAGTCCCATGAGGATTCAGAGAGAGAGGAGGAGGAGTGTGATGCTTACTT 1773
 Db 1768 CCTTGGAGTCCCATGAGGATTCAGAGAGAGAGGAGGAGGAGTGTGATGCTTACTT 1826
 Qy 1774 ATGAGCTTGACCAT-ACAGCTATCGGTATTCAGAAATATGAAACAAATCTCTG-ACAAA 1831
 Db 1827 ATGAGCTTGACCATACAGCTATCGGTATTCAGAAATATGAAACAAATCTCTGACAAA 1886
 Qy 1832 AGACCAAGCTTTAAGTTTCACAGAGTGCCTGGGCTTGTGATGATATCA-TTCCCTTTG 1890
 Db 1887 AGACCAAGCTTTAAGTTTCACAGAGTGCCTGGGCTTGTGATGATATCA-TTCCCTTTG 1946
 Qy 1891 CATTTCCCATTAATGAAACCTTTGACCTGTGAACCTTGCATCTGTATTAATACTAAA 1950
 Db 1947 CATTTCCCATTAATGAAACCTTTGACCTGTGAACCTTGCATCTGTATTAATACTAAA 2005
 Qy 1951 TTCCCAATTAAG 1962
 Db 2006 TTCCCAATTAAG 2017
 RESULT 12
 AF149783 1333 bp. mRNA linear PRI 02-JUL-2001
 LOCUS AF149783 Homo sapiens L-selectin ligand sulfotransferase mRNA, complete cds.
 DEFINITION
 ACCESSION AF149783
 VERSION AF149783.1 GI:13897503
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 1333)
 Yeh,J.C., Hiraoaka,N., Petryniak,B., Nakayama,J., Ellies,L.G.,
 Rubuka,D., Hindsgraul,O., March,J.D., Lowe,J.B. and Fukuda,M.
 Novel sulfated lymphocyte homing receptors and their control by a
 Core1 extension beta 1,3-N-acetylglucosaminyltransferase
 Cell 105 (7), 957-969 (2001)
 JOURNAL MEDLINE
 MEDLINE 2133252
 PUBMED 11439191
 REFERENCE 2 (bases 1 to 1333)
 Hiraoaka,N. and Fukuda,M.
 DIRECT SUBMISSION
 TITLE Submitted (10-MAY-1999) Glycobiology, The Burnham Institute, 10901
 JOURNAL North Torrey Pines Road, La Jolla, CA 92037, USA
 FEATURES
 source 1..1333

Db	1947	CATTITCCCATGACATGAGACGCTTTCAGCTGTAAAGC-TGCCATCTGTATATCTAATAA	2005
QY	1951	TTCCCAATAAG 1962	
Db	2006	TTCCCAATAAG 2017	
RESULT 11			
LOCUS	AF131235		
DEFINITION	Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA,		
ACCESSION	AF131235		
VERSION	AF131235.1		
KEYWORDS	GI:4927113		
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 2032)		
JOURNAL	Blstrup, A., Bhakta, S., Lee, J. K., Below, Y. Y., Gunn, M. D., Zuo, F. R.,		
MEDLINE	Huang, C. C., Kannagi, R., Kosen, S. D. and Hemmerich, S.		
PUBMED	Sulfotransferases of two specificities function in the		
REFERENCE	reconstitution of high endothelial cell ligands for L-selectin		
AUTHORS	J. Cell Biol. 145 (4), 899-910 (1999)		
TITLE	9264336		
JOURNAL	1030415		
MEDLINE	2 (bases 1 to 2032)		
PUBMED	Blstrup, A., Tangemann, K., Bhakta, S., Lee, J. K., Below, Y. Y.,		
REFERENCE	Gunn, M. D., Zuo, F. R., Huang, C. C., Kannagi, R., Kosen, S. D. and		
AUTHORS	Hemmerich, S.		
TITLE	Direct Submision		
JOURNAL	Submitted (25-FEB-1999) Respiratory Diseases, Roche Bioscience,		
MEDLINE	3401 Hillview Avenue, Palo Alto, CA 94304, USA		
PUBMED	Location/Qualifiers		
REFERENCE	1. 2032		
AUTHORS	/organism="Homo sapiens"		
TITLE	/mol_type="mRNA"		
JOURNAL	/db_xref="taxon:9606"		
MEDLINE	/cell_type="high endothelial"		
PUBMED	/tissue_type="tonsil"		
REFERENCE	182, 1342		
AUTHORS	/function="biosynthesis of endothelial ligands for		
TITLE	L-selectin		
JOURNAL	/note="HEC-G1cNAc-6-ST"		
MEDLINE	/codon_start=1		
PUBMED	/product="N-acetylglucosamine 6-O-sulfotransferase"		
REFERENCE	/protein_id="AA033015.1"		
AUTHORS	/db_xref="GI:4927114"		
TITLE	/translation="MLPKKKKLLFLVSOAIIALFFHYSIISSLSKAQPERNH		
JOURNAL	VLIVSRGSSSEFGQLFEGHPDVFYLMPEAWHWYFKOSTAWMLMAVRDLIRAFV		
MEDLINE	LCOMSVDAWMEPPROSSLFQWENSRALCSAPACDIIIPDELIIPAHORLCSQOP		
PUBMED	FEVKEKASRSHVILVLEVEFFNLQSLYPLIKPSTLMLIVHLVDRPRAVERSERBK		
REFERENCE	GDMILDSRIWNGHEOKLKKEDOPYVYMWYICQSLIYTIQSLPAKRLERYLVRY		
AUTHORS	EDLARAVAOISRMYEFVGLFELPHLOTWHTNIRGKGMDPHAFHTNARALNYSQAM		
TITLE	RMSLPYKYSRLQACGDAMNLIGYRRVRSBQERNLLDLSTWYPEQIH"		
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			

```

Db 1228 CTGTGGGATGCAATGATTTGCTGGGCTACCGCCACGATGATGACAGAACAGAG 1287
QY 1234 AAACCTGTTCTGGATCTTCTGTCTACCTGAGACGTCCTGAGCAATTCACATAGAGG 1293
Db 1288 AAACCTGTTCTGGATCTTCTGTCTACCTGAGACGTCCTGAGCAATTCACATAGAGG 1347
QY 1294 TTGAGAGGCTTGTGCTGACACGTCGTCAGCTGACCTGACCTGTCCTGATGCTTGA 1353
Db 1348 TTGAGAGGCTTGTGCTGACACGTCGTCAGCTGACCTGACCTGTCCTGATGCTTGA 1407
QY 1354 GCCTTGCTGATCTGTCGACCTTAACTGATGCTGCTGCTGATGACACGTCGATG 1413
Db 1408 GCTTGCTGATCTGTCGACCTTAACTGATGCTGCTGCTGATGACACGTCGATG 1467
QY 1414 TGTGTCCACAGCTGTCACAGAGGACTTTTGTGTCAGTCTGCTGCTGATGACACAG 1473
Db 1468 TGTGTCCACAGCTGTCACAGAGGACTTTTGTGTCAGTCTGCTGCTGATGACACAG 1527
QY 1474 ACTGGGACCTTATGTCGACGACACATCCACAGTGAACAGGATTTGCTTCTTC 1533
Db 1528 ACTGGGACCTTATGTCGACGACACATCCACAGTGAACAGGATTTGCTTCTTC 1587
QY 1534 TTTTCTGATCTCTCTGTCGTCGACGTCGACGTCGACGTCGTCGTCGTCGTCGTCG 1593
Db 1588 TTTTCTGATCTCTCTGTCGTCGACGTCGACGTCGACGTCGTCGTCGTCGTCGTCG 1647
QY 1594 GCACGACACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1653
Db 1648 GCACGACACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1707
QY 1654 GGGAAATGATCTTTCACCAAGAGCTGACAGCATTTTCCACAGAGATGCGAATTCGAG 1713
Db 1708 GGGAAATGATCTTTCACCAAGAGCTGACAGCATTTTCCACAGAGATGCGAATTCGAG 1767
QY 1714 CCTTGGAGTTCCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1773
Db 1768 CCTTGGAGTTCCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1826
QY 1774 ATGAGCTTGACAT-ACAGCTATGCTGATGATGATGATGATGATGATGATGATGATG 1831
Db 1827 ATGAGCTTGACATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1886
QY 1832 AGAGCAAGCTTCTTACAGTTCACAGGTCCTGCTGCTGATGATGATGATGATGATGATG 1880
Db 1887 AGAGCAAGCTTCTTACAGTTCACAGGTCCTGCTGCTGATGATGATGATGATGATGATG 1946
QY 1891 CATTTTCCATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1950
Db 1947 CATTTTCCATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2005
Y 1951 TTCCCAATTAAG 1962
Db 2006 TTCCCAATTAAG 2017

RESULT 10
BD134772 2032 bp DNA linear PAT 18-SEP-2002
LOCUS Glycosylsulfoltransferase-3.
DEFINITION BD134772
ACCESSION BD134772
VERSION BD134772.1 GI:23229717
KEYWORDS JP 2002507409 A/1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2032)
AUTHORS Bistup,A., Rosen,S.D., Tangemann,K. and Hemmerlich,S.
TITLE Glycosylsulfoltransferase-3
JOURNAL Patent: JP 2002507409-A 12-MAR-2002;
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA, SYNTAX INC
OS Homo sapiens (human)
PN JP 2002507409-A/1

```

```

PD 12-MAR-2002
PF 26-FEB-1999 JP 2000537979
PR 20-MAR-1998 US 09/045284,12-NOV-1998 US 09/190911 PI
ANNEETTE BISTRUP, STEVEN D ROSEN, KIRSTEN TANGEMANN, STEFAN PI
HEMERICH
PC C12N9/10,A01K67/027,A61K45/00,A61K48/00,A61P29/00,A61P37/06,
PC A61P43/00,
PC C12N1/15,C12N1/21,C12N5/10,C12N15/09,C12Q1/48,C12N5/00,C12N15/
PC 00
CC Glycosylsulfoltransferase-3
FH Key Location/Qualifiers
FT Source 1..2032
FT 1..2032
FEATURES
source 1..2032
location/Qualifiers
BASE COUNT 468 a 569 c 490 g 505 t
ORIGIN
Query Match 87.4%; Score 1784.8; DB 6; Length 2032;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1850; Conservative 0; Mismatches 17; Indels 5; Gaps 5;
QY 94 AAGCCCAAGCCACAAGGCTTCCACTTCCACGACACATGCTACTGCTAAAAAATGAGCT 153
Db 148 AAGCCGCTGTTGCAAGGCTTCCACTTCCACGACACATGCTACTGCTAAAAAATGAGCT 207
QY 154 CTTGCTTTTCTGCTTTTCCACGATGCGCATTTGCTGCTATCTTCCACATGACACCA 213
Db 208 CTTGCTTTTCTGCTTTTCCACGATGCGCATTTGCTGCTATCTTCCACATGACACCA 267
QY 214 CAACATGAGCTCCCTGCTGCTATGAGGACACAGCCGAGCGATGACGTCGCTGCTGTC 273
Db 268 CAACATGAGCTCCCTGCTGCTATGAGGACACAGCCGAGCGATGACGTCGCTGCTGTC 327
QY 274 TTTCTGGGCTTGGCTCTTCTTTTGGGGGACGCTTTTGGGACACCGACCGATGTTT 333
Db 328 TTTCTGGGCTTGGCTCTTCTTTTGGGGGACGCTTTTGGGACACCGACCGATGTTT 387
QY 334 CTACCTATGAGACCCCGCTGTCGACGTCGTCGATGATGATGATGATGATGATGATGATG 393
Db 388 CTACCTATGAGACCCCGCTGTCGACGTCGTCGATGATGATGATGATGATGATGATGATG 447
QY 394 GTCGACATGAGCTGTCGTCGATGATGATGATGATGATGATGATGATGATGATGATGATG 453
Db 448 GTCGACATGAGCTGTCGTCGATGATGATGATGATGATGATGATGATGATGATGATGATG 507
QY 454 TGATGCTTACATGAGACCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 513
Db 508 TGATGCTTACATGAGACCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 567
QY 514 CCGGGCCCTGTTCTGTCGACCTGCTGTCGATGATGATGATGATGATGATGATGATGATG 573
Db 568 CCGGGCCCTGTTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 627
QY 574 GGTCTACGACGCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 633
Db 628 GGTCTACGACGCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 687
QY 634 CTCTACGACGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 693
Db 688 CTCTACGACGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 747
QY 694 GCTGCTGAAGACCCCTCTCAACCTGTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 753
Db 748 GCTGCTGAAGACCCCTCTCAACCTGTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 807
QY 754 CGTGTCCGTTCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 813
Db 808 CGTGTCCGTTCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 867

```

DB 1399 ACCCTTGGCTACATCTCTGACCTTACATATGCTGTGGGATATCACTGAGTGTAG 1458
QY 1413 TTGTGTCCACAGTGTCTCAAGCAGAGAAGACTTTTGTGTCCATGCTGTGTCTAGAAACA 1472
DB 1459 TTGTGTCCACAGTGTCTCAAGCAGAGAAGACTTTTGTGTCCATGCTGTGTCTAGAAACA 1518
QY 1473 GACTGGGGAACCTTATGTAGACACATCCACAGTGAACAGGGTATTTGCTCTTCT 1532
DB 1519 GACTGGGGAACCTTATGTAGACACATCCACAGTGAACAGGGTATTTGCTCTTCT 1578
QY 1533 CTTTCTTGTATCTCTGCTGTGGGAGTCTTCAAGACTTTTGGCTGGAGGCTATTA 1592
DB 1579 CTTTCTTGTATCTCTGCTGTGGGAGTCTTCAAGACTTTTGGCTGGAGGCTATTA 1638
QY 1593 AGCAGACACAGTATCAGTGAATGATCATTAACCTCCCTGTCCACATCTTGGCCAA 1652
DB 1639 AGCAGACACAGTATCAGTGAATGATCATTAACCTCCCTGTCCACATCTTGGCCAA 1698
QY 1653 GGGGAATGATCTTTCACCAAGAGCTCACAGCATTTTCCACAGATGCGAATTTCTGA 1712
DB 1699 GGGGAATGATCTTTCACCAAGAGCTCACAGCATTTTCCACAGATGCGAATTTCTGA 1758
QY 1713 GCCCTTGGAGTCCCAATGGGATTCAGAGAGAGGAGTGGGAACAAGTTGGATGCTACT 1772
DB 1759 GCCCTTGGAGTCCCAATGGGATTCAGAGAGAGGAGTGGGAACAAGTTGGATGCTACT 1817
QY 1773 TATGAGCTTGACCAT-ACAGTATCGTATATGAGAAATATGAACAAATCTCTG-ACAA 1830
DB 1818 TATGAGCTTGACCATACAGTATCGTATATGAGAAATATGAACAAATCTCTGACAA 1877
QY 1831 AAGAGCAAGCTTATAGTTTACCAAGAGTGGCTGGCTGTGATTTGATATCA-TTTCCTTT 1889
DB 1878 AAGAGCAAGCTTATAGTTTACCAAGAGTGGCTGGCTGTGATTTGATATCA-TTTCCTTT 1937
QY 1890 GCATTTTCCATTTACATTAAGAAACTTGCAGTGTGAACTGGCATCTGTATATATAA 1949
DB 1938 GCATTTTCCATTTACATTAAGAAACTTGCAGTGTGAACTGGCATCTGTATATAA 1996
QY 1950 ATTCCCAATTAAGGTTCTGTATAGAAATGCTCTTTTATATGCTCTTAAATTTAGCAGTA 2009
DB 1997 ATTCCCAATTAAGGTTCTGTATAGAAATGCTCTTTTATATGCTCTTAAATTTAGCAGTA 2056
QY 2010 AATGTTTCA 2018
DB 2057 AATGTTTCA 2065

RESULT 9
AR203335
LOCUS AR203335 2032 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 2 from patent US 6365365.
ACCESSION AR203335
VERSION AR203335.1 GI:21499698
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2032)
AUTHORS Bistup,A., Rosen,S.D., Tangemann,K. and Hemmerich,S.
TITLE Method of determining whether an agent modulates glycosyl
JOURNAL Patent: US 6365365-A 2 02-APR-2002;
FEATURES
source location/Qualifiers
BASE COUNT 468 a 569 c 490 g 505 t
ORIGIN
Query Match 87.4%; Score 1784.8; DB 6; Length 2032;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1850; Conservative 0; Mismatches 17; Indels 5; Gaps 5;
QY 94 AAGCCCAAGCCCAAGGCTTCCACTTACGACACAAATGCTACTAGCCCTTAATAAATGAGCT 153

DB 148 AAGCCCGCTTCCAGAGTCTTCCACTTACGACAAAGCTACTGCTTAATAAATGAGCT 207
QY 154 CCGGCTTTTCTGTTTCCAGATGGCCATCTTGGCTATCTTCCATGATGACAGCA 213
DB 208 CCGGCTTTTCTGTTTCCAGATGGCCATCTTGGCTATCTTCCATGATGACAGCA 267
QY 214 CAACATCAGCTCCCTGTCTATGAAAGGACAGCCGAGCCATGACAGTCTGTTCTGTC 273
DB 268 CAACATCAGCTCCCTGTCTATGAAAGGACAGCCGAGCCATGACAGTCTGTTCTGTC 327
QY 274 TTCTTGGGCTCTGGCTCTTCTTTTGTGGGACAGCTTTTGGGACAGCTATTTT 333
DB 328 TTCTTGGGCTCTGGCTCTTCTTTTGTGGGACAGCTTTTGGGACAGCTATTTT 387
QY 334 CTACCTGATGAGAGCCGCTGGACAGTGTGATGATCTTCAAGCAGACAGCCGCTGAT 393
DB 388 CTACCTGATGAGAGCCGCTGGACAGTGTGATGATCTTCAAGCAGACAGCCGCTGAT 447
QY 394 GCTGCATATGCTGTGCGGATCTGATACGGGCGCTTCTTGTGCGATGAGCGTCTT 453
DB 448 GCTGCATATGCTGTGCGGATCTGATACGGGCGCTTCTTGTGCGATGAGCGTCTT 507
QY 454 TGATGCTTACATGGAACCTGTGCTCCGAGACAGTCCAGCTTTCATGTTGAGAGAC 513
DB 508 TGATGCTTACATGGAACCTGTGCTCCGAGACAGTCCAGCTTTCATGTTGAGAGAC 567
QY 514 CCGGCGCTGTGTTTGCACCTGCTGTGACATCATCCCAAGATGAATATCCCGCG 573
DB 568 CCGGCGCTGTGTTTGCACCTGCTGTGACATCATCCCAAGATGAATATCCCGCG 627
QY 574 GGTCTACAGTCTGCTCTGTGACAGTCAAGCCCTTGAAGTGTGAGAGAGCGTCCG 633
DB 628 GGTCTACAGTCTGCTCTGTGACAGTCAAGCCCTTGAAGTGTGAGAGAGCGTCCG 687
QY 634 CTCTACAGTCTGCTCTGTGACAGTCAAGCCCTTGAAGTGTGAGAGAGCGTCCG 693
DB 688 CTCTACAGTCTGCTCTGTGACAGTCAAGCCCTTGAAGTGTGAGAGAGCGTCCG 747
QY 694 GGTCTACAGTCTGCTCTGTGACAGTCAAGCCCTTGAAGTGTGAGAGAGCGTCCG 753
DB 748 GGTCTACAGTCTGCTCTGTGACAGTCAAGCCCTTGAAGTGTGAGAGAGCGTCCG 807
QY 754 GGTCTACAGTCTGCTCTGTGACAGTCAAGCCCTTGAAGTGTGAGAGAGCGTCCG 813
DB 808 GGTCTACAGTCTGCTCTGTGACAGTCAAGCCCTTGAAGTGTGAGAGAGCGTCCG 867
QY 814 GGTCTACAGTCTGCTCTGTGACAGTCAAGCCCTTGAAGTGTGAGAGAGCGTCCG 873
DB 868 GGTCTACAGTCTGCTCTGTGACAGTCAAGCCCTTGAAGTGTGAGAGAGCGTCCG 927
QY 874 GGTCTACAGTCTGCTCTGTGACAGTCAAGCCCTTGAAGTGTGAGAGAGCGTCCG 933
DB 928 GGTCTACAGTCTGCTCTGTGACAGTCAAGCCCTTGAAGTGTGAGAGAGCGTCCG 987
QY 934 GGTCTACAGTCTGCTCTGTGACAGTCAAGCCCTTGAAGTGTGAGAGAGCGTCCG 993
DB 988 GGTCTACAGTCTGCTCTGTGACAGTCAAGCCCTTGAAGTGTGAGAGAGCGTCCG 1047
QY 994 AATGTTTCAATCTGTGAGATTTGGAATTTTGGCCCATCTTCAAGACTGGGTGATTAAT 1053
DB 1048 AATGTTTCAATCTGTGAGATTTGGAATTTTGGCCCATCTTCAAGACTGGGTGATTAAT 1107
QY 1054 CACCCGAGCAGAGGAGTGTGACAGTCTTCCACAAATGCCAGGATGCCCTTAA 1113
DB 1108 CACCCGAGCAGAGGAGTGTGACAGTCTTCCACAAATGCCAGGATGCCCTTAA 1167
QY 1114 TGTCTCCAGGCTTGGCTGTGTTTGGCTTATGAAAGTTTCTGACATTCAGAAC 1173
DB 1168 TGTCTCCAGGCTTGGCTGTGTTTGGCTTATGAAAGTTTCTGACATTCAGAAC 1227
QY 1174 TGTCTCCAGGCTTGGCTGTGTTTGGCTTATGAAAGTTTCTGACATTCAGAAC 1233

Db 123928 CCGTGTGGGAGACCTTCAGACACTTTGTGGCCCTGGAGCCCTATTAGACAGACAGTA 123869
Qy 1607 TCAGTGAATGATCATTAACCTCCCTGTCACACTTTGGCCAAATGGGGAATGATCT 1666
Db 123868 TCAGTGAATGATCATTAACCTCCCTGTCACACTTTGGCCAAATGGGGAATGATCT 123809
Qy 1667 TCACAAAGAGCTCACAGACTTTTCCACAGAGATGCGAATTTGTAGCCCTTGGAGTCC 1726
Db 123808 TCACAAAGAGCTCACAGACTTTTCCACAGAGATGCGAATTTGTAGCCCTTGGAGTCC 123749
Qy 1727 CAATGGGATTCAGAGGAAGAGTGGGAACAGGTTGGATGCCCTACTTTAGAGCTTGACCA 1786
Db 123748 CAGT-GGATTCAGAGGAAGAGTGGGAACAGGTTGGATGCCCTACTTTAGAGCTTGACCA 123690
Qy 1787 T-ACAGTATCGGTATATGAGAAATATGAAACAAATCTCTG-ACAAAGAGCAAGCTCTT 1844
Db 123689 TCACAGATATCGGTATATGAGAAATATGAAACAAATCTCTGACAAAGAGCAAGCTCTT 123630
Qy 1845 AAGTTCACAGAGTGGCTGGCTGATTTGAATATCA-TTCCCTTTGCAATTTCCCATTA 1903
Db 123629 AAGTTCACAGAGTGGCTGGCTGATTTGAATATCAATCCCTCTGATTTTCCCATTA 123570
Qy 1904 CATGAAATCTTGAACCTGTAACCTTGCATCTGTATTAATTAATTAATTTCCCAATAGG 1963
Db 123569 CATGAAATCTTGAACCTGTAACCTTGCATCTGTATTAATTAATTAATTTCCCAATAGG 123511
Qy 1964 TTCTGTATGAGATGCTCCCTTTTATAGCTTTTAAATTAATTAATTAATTAATTTTAA 2023
Db 123510 TTCTGTATGAGATGCTCCCTTTTATAGCTTTTAAATTAATTAATTAATTAATTTTAA 123451
Qy 2024 TGGGATCCTAAAAAAA 2042
Db 123450 TGGGATCCTAAGCTAAGAA 123432

RESULT 8
AR310485
LOCUS AR310485 2065 bp DNA linear PAT 12-JUN-2003
DEFINITION Sequence 26 from patent US 6558935.
ACCESSION AR310485
VERSION AR310485.1 GI:31703448
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2065)
AUTHORS Tang,X.T., Corley,N.C., Guegler,K.J., Baughn,M.R., Lal,P., Yue,H.,
Hillman,J.L. and Azimzal,Y.
TITLE Human transferrase proteins
JOURNAL Patent: US 6558935-A 26 06-MAY-2003;
FEATURES
source 1..2065
/organism="unknown"
BASE COUNT 466 a 575 c 491 g 533 t
ORIGIN

Query Match 88.7%; Score 1813; DB 6; Length 2065;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1902; Conservative 0; Mismatches 20; Indels 7; Gaps 7;

Qy 94 AAGCCCAAGCACAAGGCTTCCACTTTCAGCAATAGTACTGCTTAAAAAATGAGACT 153
Db 140 AAGCCCAAGCACAAGGCTTCCACTTTCAGCAATAGTACTGCTTAAAAAATGAGACT 199
Qy 154 CCTGCTGTTCTGTTTCCAGATGGCCATCTTGGCTTATTTCCCATGATGACAGCA 213
Db 200 CCTGCTGTTCTGTTTCCAGATGGCCATCTTGGCTTATTTCCCATGATGACAGCA 259
Qy 214 CAACATGAGTCCCTGTCTATGAGAGCAGACCCGAGCGCATGACGTGTTCTGTC 273
Db 260 CAACATGAGTCCCTGTCTATGAGAGCAGACCCGAGCGCATGACGTGTTCTGTC 319
Qy 274 TTCTGGGCGCTGTGGCTCTTCTTTTGTGGGCGAGCTTTTGGGAGACACCCAGATGTTT 333

Db 320 TTCTGGGCGCTGTGGCTCTTCTTTTGTGGGCGAGCTTTTGGGAGACCCAGATGTTT 379
Qy 334 CTACCTGATGAGACCCGCTTGGACGTGTGATGACCTTCAAGACGAGACCGCGCTGAT 393
Db 380 CTACCTGATGAGACCCGCTTGGACGTGTGATGACCTTCAAGACGAGACCGCGCTGAT 439
Qy 394 GCTGCAATGAGCTGTGGGAGATCTGATACGGGCGCTTCTTGTGTCGATGAGCGTCT 453
Db 440 GCTGCAATGAGCTGTGGGAGATCTGATACGGGCGCTTCTTGTGTCGATGAGCGTCT 499
Qy 454 TGATGCTACATGAGAACTGTCCCGGAGACAGTCAGACCTTTTTCAGTGGGAGACAG 513
Db 500 TGATGCTACATGAGAACTGTCCCGGAGAGACAGTCAGACCTTTTTCAGTGGGAGACAG 559
Qy 514 CCGGCGCTGTGTTTGTGACCTGCTGTGACATCATCCCAAGATGAAATATCTCCCG 573
Db 560 CCGGCGCTGTGTTTGTGACCTGCTGTGACATCATCCCAAGATGAAATATCTCCCG 618
Qy 574 GCTGCTACAGAGCTCTGTGACAGTCAACAGCCCTTTGAGGTGG-TGGAGAAAGGCTGCC 632
Db 619 GCTGCTACAGAGCTCTGTGACAGTCAACAGCCCTTTGAGGTGGAGAAAGGCTGCC 678
Qy 633 GCTGCTACAGAGCTGTGCTCAAGGAGTGGCGCTTCTCAACCTGACGTCCCTAC 692
Db 679 GCTGCTACAGAGCTGTGCTCAAGGAGTGGCGCTTCTCAACCTGACGTCCCTAC 738
Qy 693 CGCTGCTAAGAGCCCTCCCTCAACCTGACATGCTGACCTGCTCCGGGAGACCCCGGG 752
Db 739 CGCTGCTAAGAGCCCTCCCTCAACCTGACATGCTGACCTGCTCCGGGAGACCCCGGG 798
Qy 753 CGCTGCTCCGTTCCCGAAGAGCAGCAAGGAGATCTGATGATGACAGTGGCATTTGTA 812
Db 799 CGCTGCTCCGTTCCCGAAGAGCAGCAAGGAGATCTGATGATGACAGTGGCATTTGTA 858
Qy 813 TGGGCGACATGAGCAAAACTCAAGAGAGAGACCAACCTTACTATGTATGACGTCA 872
Db 859 TGGGCGACATGAGCAAAACTCAAGAGAGAGAGACCAACCTTACTATGTATGACGTCA 918
Qy 873 TCTGCAAGAGCAGTGGAGATCTCAAGACATCAATCTGCGCAAGGCGCTGGCAGG 932
Db 919 TCTGCAAGAGCAGTGGAGATCTCAAGACATCAATCTGCGCAAGGCGCTGGCAGG 978
Qy 933 AAGCTTACCTGCTGTGCTATGAGACCTGTGCTGAGCCCTGTGGCCAGACTTCC 992
Db 979 AAGCTTACCTGCTGTGCTATGAGACCTGTGCTGAGCCCTGTGGCCAGACTTCC 1038
Qy 993 GAATGATGAATTTGGGATTTGGAATTTGCCCATCTTCAAGACCTGGGTGCTATACA 1052
Db 1039 GAATGATGAATTTGGGATTTGGAATTTGCCCATCTTCAAGACCTGGGTGCTATACA 1098
Qy 1053 TCACCCGAGGCAAGGCGATGGGTGACACAGCTTTCCACACAATAGCCAGGAGTCCCTTA 1112
Db 1099 TCACCCGAGGCAAGGCGATGGGTGACACAGCTTTCCACACAATAGCCAGGATGCCCTTA 1158
Qy 1113 ATGCTCCAGGCTTGGCGCTGTCTTTGCCCTATGAAGAGTTTCTGACTTGAAG 1172
Db 1159 ATGCTCCAGGCTTGGCGCTGTCTTTGCCCTATGAAGAGTTTCTGACTTGAAG 1218
Qy 1173 CCTGTGGGAGTCCATGATTTGCTGGGCTACCGCAGCTGATCTGAATCAAGACAGA 1232
Db 1219 CCTGTGGGAGTCCATGATTTGCTGGGCTACCGCAGCTGATCTGAATCAAGACAGA 1278
Qy 1233 GAAACCTGTGCTGATCTTCTGCTTACCTGAGCTGCTGAGAGAAATTCAGTAAGG 1292
Db 1279 GAAACCTGTGCTGATCTTCTGCTTACCTGAGCTGCTGAGAGAAATTCAGTAAGG 1338
Qy 1293 GTTGAAGAGCTTTGCTCCACCTGTGTGCTCAGCTCACTTCTCTGATGCTTCTG 1352
Db 1339 GTTGAAGAGCTTTGCTCCACCTGTGTGCTCAGCTCACTTCTCTGATGCTTCTG 1398
Qy 1353 AGCTTGGCTACATCTGTGAGCTTAACTACATGCTGTGGGTATCACATGATGTGAG 1412

THIS PAGE BLANK (USPTO)

Db	1172	GCCATTCATCTTCGCTAGGAATGCGGCAACGTCCTCCAGGCTGCGGCACGCGFTG	1231
Qy	1142	CCCTATGAAAGGTTTCTCGACTTCAGAAAGCCTGTGGGATGCCATGATTTGCTGGGC	1201
Db	1232	CCCTTCACTAGATCTGCGGCTGAGAGGTGTGCGCGGCGCTGCAGCTGCTGGGC	1291
Qy	1202	TACCGCCACGTCAGATCTGAACAAGAACGAGAAACCTGTGCTGATCTTCTG	1255
Db	1292	TACCGGCTGTGTACTCTGCGGACGAGCGGTGACCTCACCTGATCTGTG	1345

Search completed: August 14, 2003, 11:38:20
 Job time : 514.533 secs

FT	/tag= h
FT	/cons_splice= (5'site:YES, 3'site:NO)
FT	47939..49746
FT	/tag= i
FT	/number= 5
FT	/note= "Includes 17 base pairs of 5'UTR, the ORF
FT	and all of 3'UTR"
FT	47939..47955
FT	/tag= j
FT	/note= "Portion of 5' untranslated region (5'UTR)"
FT	47956..49128
FT	/tag= k
FT	/product= "Human glycosyl transferase-4alpha
FT	(GST-4alpha)"
FT	49129..49746
FT	/tag= l
FT	83257..83347
FT	/tag= m
FT	/label= 4a_5U2
FT	83348..96412
FT	/tag= n
FT	/cons_splice= (5'site:NO, 3'site:NO)
FT	96413..96484
FT	/tag= o
FT	/label= 4a_5U1
FT	96485..98456
FT	/tag= p
FT	/cons_splice= (5'site:NO, 3'site:NO)
FT	98457..99968
FT	/tag= q
FT	/note= "Includes 17 base pairs of 5'UTR, the ORF
FT	and all of 3'UTR"
FT	98457..98473
FT	/tag= r
FT	/note= "Portion of 5' untranslated region (5'UTR)"
FT	98474..99661
FT	/tag= s
FT	/product= "Human glycosyl transferase-4beta
FT	(GST-4beta)"
FT	99662..99968
FT	/tag= t
PN	WO200106015-A1.
PD	25-JAN-2001.
XX	
XX	19-JUL-2000; 2000OWO-US19741.
PF	
XX	20-JUL-1999; 99US-0144694.
PR	13-JUL-2000; 2000OUS-0593826.
XX	
PA	(REGC) UNIV CALIFORNIA.
PI	Rosen SD, Lee JK, Hemmerich S;
XX	
DR	WPI: 2001-138471/14.
DR	P-PSDB: AAY72639, AAY72640.
XX	
PT	New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT	diagnostic and therapeutic agent screening applications -
XX	
PS	Example 1; Page 62-104; 128pp; English.
XX	
CC	The present sequence is human glycosyl sulfotransferase-4 (GSR-4) genomic
CC	DNA encoding GSR-4alpha and GSR-4beta. GSR-4 gene is found on
CC	chromosome 16q23.1.
CC	GSR is a type 2 membrane protein useful for inhibiting a binding event
CC	between a selectin and a selectin ligand, which comprises contacting the
CC	selectin with a non-sulphated selectin ligand, GSR and a small molecular
CC	agent that inhibits the sulphation activity of GSR. GSR is also useful
CC	in inhibiting a selectin mediated binding event. GSR is useful in gene
CC	therapy to treat disorders such as acute or chronic inflammation,
CC	systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis


```

FT  /tag- c
XX  W0200106015-A1.
XX  25-JAN-2001.
XX  19-JUL-2000; 2000MO-US19741.
XX  20-JUL-1999; 99US-0144694.
XX  13-JUL-2000; 2000US-0593828.
XX  (REGC ) UNIV CALIFORNTA.
XX  PI Rosen SD, Lee JK, Hemmerich S;
XX  WPI: 2001-138471/14.
XX  P-PSDB: AAY72640.
XX  New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX  diagnostic and therapeutic agent screening applications
XX  Claim 6; Fig 4A; 128pp; English.
XX  The present sequence is human glycosyl sulfotransferase-4beta (GST-4
XX  beta) cDNA. GST-4 gene is found on chromosome 16q23.1.
XX  GST is a type 2 membrane protein useful for inhibiting a binding event
XX  between a selectin and a selectin ligand, which comprises contacting the
XX  selectin with a non-sulphated selectin ligand, GST and a small molecular
XX  agent that inhibits the sulphation activity of GST. GST is also useful
XX  in inhibiting a selectin mediated binding event. GST is useful in gene
XX  therapy to treat disorders such as acute or chronic inflammation,
XX  systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
XX  nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
XX  glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
XX  disease, Grave's disease, adenitis, hypoparathyroidism, periculous
XX  anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
XX  dermatitis, myocarditis, regional enteritis, adult respiratory distress
XX  syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
XX  bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
XX  during transplantation.
XX  Sequence 1694 BP; 259 A; 586 C; 514 G; 335 T; 0 other;
SQ
Query Match 20.2%; Score 412.8; DB 22; Length 1694;
Best Local Similarity 64.4%; Pred. No. 1.2e-110;
Matches 653; Conservative 0; Mismatches 352; Indels 9; Gaps 2;
QY 248 GAGGCGATGACGAGTGTCTGTCTCTCTGCGGCGCTGCTCTTTGTGGGCGAG 307
DB 303 GCGGCGGTGACGTGTGTGTGTCTCTCTGCGGCGCTGCTCTCTGTGGGCGAA 362
QY 308 CTTTTTGGGAGCACCCAGATGTTTCTACCTGATGAGCCGCGCTGCGACGTGTGATG 367
DB 363 CTCTTCAACGACGACCCCGAGGTCTTACTATATGAGCCGCGGTGCGACGTGTGAC 422
QY 368 ACCTTCAAGACAGCACCGCTGATCTGCACATGCGCTGTGCGGATCTGATACGGGCC 427
DB 423 ACCCTGTGCGAGGCGACGCGCAACGCTGACATGTGCTGTGCGACGTGTGCGCTCC 482
QY 428 GTCTCTTGTGGCAGATGAGCGTCTTGTGATGCTTACATGGAACCTGTGCTCCCGAGACAG 487
DB 483 GTCTCTCTGTGCGACATGAGAGGTGTGTGATGCTTATGTG---CCTTGGCGCGCAACCTG 539
QY 488 TCCAGCCTCTTTCATGGGAGAACAGCCGGCCCTGTGTTCTGTGACCTGACCTGTGACATC 547
DB 540 TCCGACCTCTTTCATGGGCGCGGTGAGCCGTGACATGTGTGTGCGACCCGCGCTGAGAGCC 599
QY 548 ATCCACACAGATGAATCATCTCCCGGGCTCATCTGCAAGGCTCTGTGCAATCAACAGCC 607
DB 600 TTTCCTCCGAGGCGCATCACAGAGGCGGTGTGCAAGCACTGTGCGCGCGGACAGTCC 659
QY 608 TTGTAGGTGTGAGAGGCGCTGCTCTACAGCCACGTGTGTCTCAAGAGAGTGTGCC 667

```

```

DB 660 TTACACCTGAGCCCGGAGGCGCTGCGCTCTACAGCCACGTGTGTCTCAAGAGGTGCC 719
QY 668 TTCTTCAACCTGACGTCCCTTACCCGTGTGTAAGACCCCTCCCTCAACCTCATATC 727
DB 720 TTCTTCAACCTGACGTGTCTTACCCGTGTGTAAGACCCCTCCCTCAACCTCATATC 779
QY 728 GTGCACTGTGTCCGAGCCCGGCGCTGTTCCTGTTCCGTTCCGAGACGACCAAGGAGAT 787
DB 780 GTGCACTGTGTCCGAGCCCGGCGCTGTGTGCGTGTGCGGAGACGACCAAGGAGCT 839
QY 788 CTCAATGATGACATGCTGTGTGTGTGAGGCGACATGACAAAATCTCAAGAGAGAC 847
DB 840 CTGCGCGGTGACCAAGGATCTGTGTGTGTGACCAAGGCGACGTTGAGGCGAGCC 899
QY 848 CAACCTACTATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 907
DB 900 GCGCTGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 959
QY 908 CAGTCTTGTCCCAAGGCGCTGACAGAAAGCTTACCTGTGTGTGTGTGTGTGTGTGTGTGT 967
DB 960 CTCAAGGCGCGACCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1019
QY 968 CGAGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1027
DB 1020 CGGAGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1079
QY 1028 CATCTTACAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1081
DB 1080 CAGTCTGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1139
QY 1082 GCTTTCACACCAAAATGCCAGGATGCCCTTATGTCTTCCAGGCTTGTGTGTGTGTGTGT 1141
DB 1140 GCTTTCACCAAAATGCCAGGATGCCCTTATGTCTTCCAGGCTTGTGTGTGTGTGTGTGT 1199
QY 1142 CCCATTAAGAAATTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1201
DB 1200 CCGTTTGCAGATGTCGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1259
QY 1202 TACCGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1265
DB 1260 TACCGCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1313

```

RESULT 13
ABN89506
ID ABN89506 standard; cDNA: 2544 BP.

XX AC ABN89506;
XX DT 05-SEP-2002 (first entry)
XX DE Human corneal N-acetylglucosamine-6-sulfotransferase cDNA seq ID NO:1.
XX KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme: GLCNAC6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological; chromosome 16q22; gene: ss.
XX OS Homo sapiens.
XX XX
XX FH key Location/Qualifiers
XX FT 693..1880
XX FT /tag- a
XX FT /product= "N-acetylglucosamine-6-sulfotransferase"
XX PD US2002061562-A1.
XX PD 23-MAY-2002.
XX PF 09-AUG-2001; 2001US-0927602.
XX PF 11-AUG-2000; 2000US-325773P.
XX PA (FUKU/) FUKUDA M N.

DR P-PSDB; AAE15438.
 XX Novel isolated human drug metabolizing enzymes referred as DME 1-10,
 PT useful for diagnosing, treating, or preventing disorders associated
 PT with aberrant expression of DME such as allergy, anemia, asthma,
 PT infertility
 PS Claim 5; Page 139; 143pp; English.
 XX
 XX The invention relates to human drug metabolizing enzymes referred as
 CC DME and nucleic acid molecules encoding such enzymes. Polynucleotides
 CC of the invention are useful for assessing toxicity of test compounds
 CC and in gene therapy. Sequences of the invention are useful in the
 CC diagnosis, prevention and treatment of autoimmune/inflammatory
 CC disorders such as acquired immune deficiency syndrome (AIDS), adult
 CC respiratory distress syndrome, allergies, anemia, atherosclerosis,
 CC asthma, autoimmune hemolytic anemia, contact dermatitis, Crohn's
 CC disease, glomerulonephritis, Goodpasture's syndrome, gout, Grave's
 CC disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple
 CC sclerosis, osteoarthritis, osteoporosis, psoriasis, systemic lupus
 CC erythematosus, rheumatoid arthritis, scleroderma, ulcerative colitis,
 CC uveitis, viral, bacterial, fungal, parasitic, protozoal, helminthic
 CC infections; cell proliferative disorders such as actinic keratosis,
 CC arteriosclerosis, atherosclerosis, Duchenne and Becker dystrophy,
 CC hepatitis, Cushing's syndrome, cancers, myelodysplastic syndrome,
 CC epilepsy; endocrine disorders such as disorders of the hypothalamus
 CC and pituitary resulting from lesions such as primary brain tumors,
 CC adenomas, infarction associated with pregnancy, aneurysms, vascular
 CC malformations; eye disorders such as conjunctivitis, iritis, retinitis,
 CC glaucoma, pigmentosa; metabolic disorders such as Addison's disease,
 CC cystic fibrosis, diabetes, goitre, glycogen storage diseases,
 CC hypercholesterolemia, hyperthyroidism, hypoglycaemia, lipid myopathies,
 CC Menkes syndrome, mannosidosis, obesity; gastrointestinal disorders such
 CC as dysphagia, gastric carcinoma, anorexia, nausea, gastroenteritis,
 CC hyperbilirubinemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's
 CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental
 CC disorders. The present sequence is human DME-5 protein cDNA.
 CC
 XX Sequence 1647 BP; 287 A; 560 C; 501 G; 299 T; 0 other;
 SQ
 Query Match 20.2%; Score 412.8; DB 24; Length 1647;
 Best Local Similarity 64.4%; Pred. No. 1.2e-110;
 Matches 653; Conservative 0; Mismatches 352; Indels 9; Gaps 2;

DB 532 TTTCTTCAACCTGAGTGTCTTACCCGCTGCTCAGGACCCCGCTCAACCTACGCAAC
 QY 728 GGGACCTGCTGGGAGCCCGGCGGTTCCGCGGAGAACGCAAGGAGAT 787
 DB 592 GTGACCTGTGTGGAGCCCGGCGGCTGTGCTCCGGGAGCAGACGCAAGGCT 651
 QY 788 CTCATGATTTAGACAGTGTGTGATGGGAGCATAGACCAAAACTCAAGAGAGAC 847
 DB 652 CTGGGCGGTGACAGGACATGCTGTGGGACACAGGACGTGGGTGAGGCGGACCC 711
 QY 848 CAACCTTACTGTATGTAGAGTCACTGTGCAAGCAGCTGAGATCTCAAGACATC 907
 DB 712 GGGCTGCGGTGTGTGGAGGTGTCCTGACCAAGTACGATCCGACGCGCGCA 771
 QY 908 CATGCTTCCCAAGCCCTGCGAGAACCTACTGCTGTGCTATGAGACCTGGCT 967
 DB 772 CTCACGCCGCAACCTTTGTGCGGCGGCGCTACCGCTGTGCTGTGCAAGACCTGGG 831
 QY 968 CGAGCCCTGTGGCCAGACTTCCGATGTATGAATTCGTGGATTTGAATTTGCC 1027
 DB 832 CGGAGCCGCTGCAAGAAATCCGTGCGCTACGCTTACAGTGCATGCTACAGCCA 891
 QY 1028 CATCTTCAACCTGGGTGATATACATACCCAGGCAAGGCAATGGTG-----ACCA 1081
 DB 892 CAGCTTGAGGCGCTGATCATTAACATCCAGGATCTGAGACCTGTCGCGCGCGCA 951
 QY 1082 GCTTTTCAACAAATCCAGGAGTCCCTTAATGTCTCCAGGCTGGGCTGTGTTG 1141
 DB 952 GCCTTCAAGACTTCGTCAGAGAAATGCGCTCAACGCTCCAGGCTGGGCGCATGGGCTG 1011
 QY 1142 CCCTATGAAGATTTCTGACTTCAGAAAGCTGTGGCGATGCCATGAAATTTCTGGCC 1201
 DB 1012 CCTTTGCCAAGTCGCGCGCGTGCAGAACTGTGCGTGTGCTGCAGCTGCGGGC 1071
 QY 1202 TACCCGACGTGATGATGAACAAGAAAGAACTGTGCTGATTTCTG 1255
 DB 1072 TACCGGCTGTGTACTGTAGAGCAGACGCAACCTGCTGATCTGTG 1125

RESULT 12
 AAD02700 standard; cDNA; 1694 BP.
 ID AAD02700;
 AC AAD02700;
 XX
 DT 02-MAY-2001 (first entry)
 XX
 DE Human glycosyl sulfotransferase-4beta (GST-4beta) cDNA.
 XX
 KW Human; glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive;
 KW therapy; selection binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; anemia;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; psoriasis;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 KW chromosome 16q23.1; ss.
 XX
 OS Homo sapiens.
 XX
 FH key
 FT 5' UTR Location/Qualifiers
 FT 9..188
 FT /*tag= a
 FT 189..1376
 FT CDS
 FT /*tag= b
 FT /product= "Human glycosyl sulfotransferase-4beta
 FT (GST-4beta)"
 FT /note= "CDS is referred as SEQ ID NO:21 in brief
 FT description of the figures (page no: 4)."
 FT 1377..1694
 FT 3' UTR


```

QY 154 CCTGCTGTTTGTGTTTCCAGATGGCATCTTGCTATCTTCACATGTACAGCA 213
DB 95 CCGTGGTCTTGTGTTTCCAGATGGCATCTTGCTATCTTCACATGTACAGCA 154
QY 214 CAACATCAGCTCCCTGTCTATGAAGCAGACCCGAGCCATGACAGTCTGTTCTGC 273
DB 155 CAACATCAGCTCCCTGTCTATGAAGCAGACCCGAGCCATGACAGTCTGTTCTGC 214
QY 274 TTTCTGGGCGCTGCTGCTCTTTTGTGGGAGAGCTTTTGGGAGCAGCCAGATGTTT 333
DB 215 TTCTGGGCGCTGCTGCTCTTTTGTGGGAGAGCTTTTGGGAGCAGCCAGATGTTT 274
QY 334 CTACCTGATGAGACCCGCTGGCAGCTGTGATGATGACCTTCAAGCAGACCCGCTGAT 393
DB 275 CTACCTGATGAGACCCGCTGGCAGCTGTGATGATGACCTTCAAGCAGACCCGCTGAT 334
QY 394 GGTGACATGCTGTGCGGAGATCTATACGGCCGCTCTTGTGTGGAGATGAGCTCTT 453
DB 335 GGTGACATGCTGTGCGGAGATCTATACGGCCGCTCTTGTGTGGAGATGAGCTCTT 394
QY 454 TGATGCTACATGGAACCTGTCTCCCGAGACAGTCCAGCTCTTCACTGGAGAGACAG 513
DB 395 TGATGCTACATGGAACCTGTCTCCCGAGACAGTCCAGCTCTTCACTGGAGAGACAG 454
QY 514 CCGGCGCCCTGTCTTGCACCTGCTGTGATCATCCACAAGATGAATATCCCCCG 573
DB 455 CCGGCGCCCTGTCTTGCACCTGCTGTGATCATCCACAAGATGAATATCCCCCG 514
QY 574 GGCTACATGAGAGCTCTGTGACATCAACAGCCCTTGAAGTGGAGAGAGCCGCGG 633
DB 515 GGCTACATGAGAGCTCTGTGACATCAACAGCCCTTGAAGTGGAGAGAGCCGCGG 574
QY 634 CTCTACAGCAGCAGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACATCCCTTACCC 693
DB 575 CTCTACAGCAGCAGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACATCCCTTACCC 634
QY 694 GGTGCTGAAGAGACCCCTCCCTCAACCTGATATCGTGACCTGTGTCGGGAGCCCGGCG 753
DB 635 GGTGCTGAAGAGACCCCTCCCTCAACCTGATATCGTGACCTGTGTCGGGAGCCCGGCG 694
QY 754 CGTGTCCGTTCCCGAGACGACAAAGAGAGATCATGATGACATGCGATTGTAT 813
DB 695 CGTGTCCGTTCCCGAGAGACGACAAAGAGAGATCATGATGACATGCGATTGTAT 754
QY 814 GGGGAGCAGTGAAGCAAACTCAAGAGAGAGAGCAACCCCTACTATGTATGACAGTCA 872
DB 755 GGGGAGCAGTGAAGCAAACTCAAGAGAGAGAGCAACCCCTACTATGTATGACAGTCA 814
QY 873 -TCTGCAAGAGCAGC-TGGAGATCTACAAAGAC 904
DB 815 TCTGGCAAGAGCAGCTTGAGAACTACAAAGAC 848

```

RESULT 7

AAZ20793 standard; DNA; 1926 BP.

AAZ20793;

08-DEC-1999 (first entry)

Mouse glycosyl sulfotransferase-3 coding sequence.

glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;

selectin binding interaction; inflammation; lymphocyte homing; mouse;

secondary lymph organ; ss.

Mus sp.

W09949018-A1.

30-SEP-1999.

```

PF 26-FEB-1999; 99MO-US04316.
XX
PR 20-MAR-1998; 98US-0045284.
PR 12-NOV-1998; 98US-0190911.
XX
PA (REGC) UNIV CALIFORNIA.
PA (SYNT) SYNTX USA INC.
XX
PI Bistrup A, Rosen SD, Tangemann K, Hemmerlich S;
XX
DR WPI, 1999-580442/49.
DR P-PSDB; AAY39919.
XX
PT Human and murine glycosyl sulfotransferase 3 and related
PS polynucleotides -
XX
PS Claim 4; Fig 3; 59pp; English.
XX
CC This sequence encodes the mouse glycosyl sulfotransferase-3 (GST-3) of
CC the invention. The nucleic acid sequences, probes and primers derived
CC from these, proteins and antibodies are useful in detecting homologues.
CC The sequences, antibodies and methods are useful in the diagnosis and
CC treatment of diseases associated with selectin binding interactions,
CC including conditions associated with or resulting from the homing of
CC leukocytes to sites of inflammation and the normal homing of lymphocytes
CC to secondary lymph organs.
XX
SQ Sequence 1926 BP; 426 A; 520 C; 484 G; 496 T; 0 other;

```

Query Match. 36.4%; Score 743.2; DB 20; Length 1926;
Best Local Similarity 76.5%; Pred. No. 4,6e-208;

Matches 938; Conservative 0; Mismatches 283; Indels 5; Gaps 2;

```

QY 111 TCTTCACCTTACAGCAATGCTACTGCTAAATAAAGTCCGCTGTTTCTGTTT 170
DB 388 TCTTCCTTCCGAGAGATGATGCTTTGAAGAAAGAGAGGCTGATGTTCTTGTT 447
QY 171 CCCAGATGGCATCTTGTGCTATCTTCCACATGTACAGCCACAACATCAGCTCCGT 230
DB 448 CCCAGATGATCGTTGAGCTCTCTTCATCATATGCTCCACAGACACCTTCCAG- 506
QY 231 CATAGAGCAGACGCCGAGCCAGTACAGTCTGTTGTTCTTCCGCGCTCGGCT 290
DB 507 --AGGAGGAGTCCAGAGAGCCCGTCAATGCTGCTGCTGCTGCTGCTGCTGCT 564
QY 507 --AGGAGGAGTCCAGAGAGCCCGTCAATGCTGCTGCTGCTGCTGCTGCTGCT 564
DB 291 CTTCCTTGTGGGAGAGCTTTTGGGAGCAGCAGCAGATGTTTCTACTGATGAGCCG 350
QY 351 CCTGTCAGTGTGATGACCTTCAAGCAGACAGCCGCTGATGTGACATGCTGTGC 410
DB 625 CCGGATGTGTGATGATGATCTTCAACGACAGCAGCAGCCTGGAAGCGACATGGCTGTC 684
QY 411 GGGATCTGATGAGGGGCGCTCTTGTGGAGATGAGCGCTTGTGATGCTCAATGAGAC 470
DB 685 GGGATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 744
QY 471 CTGTTCCCGGAGACAGTCCAGCTCTTTCAGTGGAGAAACGCGGCGCTGTGTTCTG 530
DB 745 CAGGCGCCCGGAGAAAGTCCAGAGCTTCCAGTGGAGAAAGCGGCGCTGTGCTAG 804
QY 531 CACCTGCTGTGATCATCTCCACAAGATGAATATCTCCCGGCGCTCACTGAGGCTCC 590
DB 805 CCGCTGTGTGTACTCTTCCGCGCAGAGATGAGCTGACCCCAAGCAGCTCAACCTGC 864
QY 591 TGAGCAGTCAAGAGCCCTTGAAGTGTGGAGAGAGCTGCGGCTCTCAAGCAGAGTGG 650
DB 865 TCTGCGGTACAGACGCCCTTGTGATGTGTGAGAGAGAGCTGCGGCTCTCAAGGCTGTGG 924
QY 651 TGCTCAAGAGAGTGCCTTCTTCAACCTCAGTCCCTTACCGGCTGTGAAGAGCCCT 710
DB 925 TACTCAAGAGAGTGCCTTCTTCAAGCTCAGGCGCTTATCACTACTACAGGAGCCCT 984

```

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
XX WPI: 2001-524255/58.

PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -

PS Claim 2; SEQ ID NO 263; 1380bp + sequence listing; English.

CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is the nucleotide
CC sequence of the 5'-end of a cDNA provided in the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;

Query Match 37.4%; Score 763.6; DB 22; Length 877;

Best Local Similarity 97.7%; Pred. No. 2.8e-214;

Matches 795; Conservative 0; Mismatches 16; Indels 3; Gaps 2;

QY 94 AAGCCCAAGCACAAGGCTTCCACTGACAGCAATGTACTGCTTAAAAAAGAACT 153
DB 35 AAGCCCGCTGTGAAGGCTTCCACTGACAGCAATGTACTGCTTAAAAAAGAACT 94
QY 154 CCTGCTGTTTGTGTTTCCAGATGAGCATCTTGGCTATCTTCCATGATGAGCA 213
DB 95 CCTGCTGTTTGTGTTTCCAGATGAGCATCTTGGCTATCTTCCATGATGAGCA 154
QY 214 CAACATGAGCTCCTGCTATGAAGCAGACCCGAGCGATGACGTGCTGTC 273
DB 155 CAACATGAGCTCCTGCTATGAAGCAGACCCGAGCGATGACGTGCTGTC 214
QY 274 TTCTGGGCGTGGGCTCTTCTTTGAGGAGCAGTTTGGGAGCAGCCAGATGTTT 333
DB 215 TTCTGGGCGTGGGCTCTTCTTTGAGGAGCAGTTTGGGAGCAGCCAGATGTTT 274
QY 334 CTACCTGATGAGCCCGGCTGACAGTGTGATGACCTTCAAGAGAGACCGCTGAT 393
DB 275 CTACCTGATGAGCCCGGCTGACAGTGTGATGACCTTCAAGAGAGACCGCTGAT 334
QY 394 GCTGACATGAGCTGTGGGATGATGAGGCGGCTTCTTGTGCGACATGAGGCTT 453
b 335 GCTGACATGAGCTGTGGGATGATGAGGCGGCTTCTTGTGCGACATGAGGCTT 394
QY 454 TGATGCTTGAAGTGAACCTGTTCCCGGAGACACTCCAGCTCTTTAGTGGAGAAAG 513
DB 395 TGATGCTTGAAGTGAACCTGTTCCCGGAGACACTCCAGCTCTTTAGTGGAGAAAG 454
QY 514 CCGGGCCCTGTGCTGACACTGCTGATCATCCCAAGATGAATCAATCCCGG 573
DB 455 CCGGGCCCTGTGCTGACACTGCTGATCATCCCAAGATGAATCAATCCCGG 514
QY 574 GGTGCTGAGGCTCTGTGACATCAACAGCCCTTGAAGTGTGAGAAAGGCTGCG 633
DB 515 GGTGCTGAGGCTCTGTGACATCAACAGCCCTTGAAGTGTGAGAAAGGCTGCG 574
QY 634 CTCTTACAGCAGCTGTGCTCAAGAGAGTGGCTTCTTCACTGACAGTCCCTTACCC 693
DB 575 CTCTTACAGCAGCTGTGCTCAAGAGAGTGGCTTCTTCACTGACAGTCCCTTACCC 634
QY 694 GCTGCTGAAGAGACCCCTCCCAACCTGCAATATGCTGACCTGTCGGGAGCCCGGGC 753
DB 635 GCTGCTGAAGAGACCCCTCCCAACCTGCAATATGCTGACCTGTCGGGAGCCCGGGC 694

QY 754 CGTGTCCGTCCTCCGAGAACCCACAAAGGAGATCTGATGACACTGCTGATGAT 813
DB 695 CGTGTCCGTCCTCCGAGAACCCACAAAGGAGATCTGATGATGACACTGCTGATGAT 754
QY 814 GGGCAGCATGATGACAAAACCTCAGAGAGAGACCAACCTTACTATGTGATGAGTGA- 872
DB 755 GGGCAGCATGATGACAAAACCTCAGAGAGAGACCAACCTTACTATGTGATGAGTGA- 814
QY 873 -TCTGCCAAGCAGC-TGAGATCTACAGACC 904
DB 815 TCTGCCAAGCAGCCTTGAGAGACCTACAGACC 848

RESULT 6
AAK93921

ID AAK93921 standard; cDNA; 877 BP.

AC AAK93921;

DT 06-NOV-2001 (first entry)

DE Human cDNA clone representative sequence; SEQ ID NO: 2381.

XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX Homo sapiens.

XX EP1130094-A2.

XX 05-SEP-2001.

XX 07-JUL-2000; 2000EP-0114089.

XX 08-JUL-1999; 99JP-0194486.

XX 11-JAN-2000; 2000JP-0118774.

XX 02-MAY-2000; 2000JP-0185765.

XX (HELI-) HELIX RES INST.

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX WPI: 2001-524255/58.

DR 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -

PS Example 11; SEQ ID NO 2381; 1380bp + sequence listing; English.

CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence was used as the
CC representative sequence from a human clone which was used in
CC homology searches to identify the clone.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;

Query Match 37.4%; Score 763.6; DB 22; Length 877;

Best Local Similarity 97.7%; Pred. No. 2.8e-214;

Matches 795; Conservative 0; Mismatches 16; Indels 3; Gaps 2;

QY 94 AAGCCCAAGCACAAGGCTTCCACTGACAGCAATGTACTGCTTAAAAAAGAACT 153
DB 35 AAGCCCGCTGTGAAGGCTTCCACTGACAGCAATGTACTGCTTAAAAAAGAACT 94

PT Intestinal GlcNAc 6-sulfotransferase -
 XX
 PS Claim 19; Fig 4; 98bp; English.
 XX
 CC The present invention provides a method of modifying an acceptor molecule
 CC by contacting the acceptor with an isolated
 CC beta1,3-N-acetylglucosaminyltransferase (beta1,3gnt) or an active
 CC fragment, where beta1,3gnt directs expression of a MECA-79 antigen. The
 CC invention also provides a method of treating or preventing an
 CC L-selectin-mediated condition by reducing the expression or activity of a
 CC beta1,3gnt that directs expression of a MECA-79 antigen. This can be done
 CC by administering to the subject an oligosaccharide L-selectin antagonist
 CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
 CC administering antibody material that specifically binds beta1,3gnt,
 CC and/or a beta1,3gnt antisense nucleic acid molecule. L-selectin
 CC sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.
 CC Alternatively, the expression or activity of LST-2 or its active
 CC fragment can be reduced in combination with reducing the expression or
 CC activity of beta1,3gnt. The method is useful for treating L-selectin
 CC mediated conditions such as Crohn's disease and ulcerative colitis,
 CC inflammatory disorders of the skin such as allergic contact dermatitis,
 CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
 CC sequence represents cDNA encoding human LST-2.
 XX
 XX Sequence 1333 BP; 286 A; 393 C; 336 G; 318 T; 0 other;
 SQ
 Query Match 61.8%; Score 1262; DB 24; Length 1333;
 Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Gaps 0;
 Matches 1262; Conservative 0; Indels 0;

107 AAGTCTTCACCTTACAGCAATGCTACTGCTTAAATAAGTCTGCTGTTCTG 166
 72 AAGGCTTTCACCTTACAGCAATGCTACTGCTTAAATAAGTCTGCTGTTCTG 131
 167 GTTTCACAGATGGCATCTTGCTCTATTCTTCACATGTACAGCCACAATCAGCTCC 226
 132 GTTTCACAGATGGCATCTTGCTCTATTCTTCACATGTACAGCCACAATCAGCTCC 191
 227 CTGTCTATGAGGACAGCCGAGGAGGATGACAGGCTGCTGCTGCTGCTGCTGCT 286
 192 CTGTCTATGAGGACAGCCGAGGAGGATGACAGGCTGCTGCTGCTGCTGCTGCT 251
 287 GGCTCTTCTTTTGTGGGAGCCTTTTGGGAGCAACCCAGATGTTTCTACCTGATGAG 346
 252 GGCTCTTCTTTTGTGGGAGCCTTTTGGGAGCAACCCAGATGTTTCTACCTGATGAG 311
 347 CCCGCTGCGACAGTGTGATGACCTTCAAGACAGACAGCCGCTGATGCTGACATGCT 406
 312 CCCGCTGCGACAGTGTGATGACCTTCAAGACAGACAGCCGCTGATGCTGACATGCT 371
 407 GTGGGAGATCTGATAGCGGCGCTCTTGTGAGCATGAGGCTCTTATGCTATCATG 466
 372 GTGGGAGATCTGATAGCGGCGCTCTTGTGAGCATGAGGCTCTTATGCTATCATG 431
 467 GAACCTGTGCTCCCGGAGACAGTCCAGCCTTTTCACTGGGAGAAACAGCCGCCCTGT 526
 432 GAACCTGTGCTCCCGGAGACAGTCCAGCCTTTTCACTGGGAGAAACAGCCGCCCTGT 491
 527 TGTGACCTGCTGCTGACATCATCCACAAGATGAATATATCCCGGGGCTCACTGAGG 586
 492 TGTGACCTGCTGCTGACATCATCCACAAGATGAATATATCCCGGGGCTCACTGAGG 551
 587 CTCTGTGAGTCAACAGCCCTTTGAGTGTGAGGAGAGGCTGCGCTCTACAGCCAC 646
 552 CTCTGTGAGTCAACAGCCCTTTGAGTGTGAGGAGAGGCTGCGCTCTACAGCCAC 611
 647 GTGTGTCTCAGAGAGGCTGCTTCTTCAACTGACAGTCCCTTACCCGCTGTGAAGAC 706
 612 GTGTGTCTCAGAGAGGCTGCTTCTTCAACTGACAGTCCCTTACCCGCTGTGAAGAC 671
 707 CCCTCCCTCAACCTGATATGTCGACCTGCTCCGGAGACCCCGGGGCTGTTCGTTCC 766

DB 672 CCCTCCCTCAACCTGATATGTCGACACCTGTCGCGGAGACCCCGGGCTGTTCGTTCC 731
 OY 767 CGAGAACGCAAAAGGAGATCTCATGATTGACAGTGGCATTTGTATGGGAGCATGAG 826
 DB 732 CGAGAACGCAAAAGGAGATCTCATGATTGACAGTGGCATTTGTATGGGAGCATGAG 791
 OY 827 CAAAACCTCAGAGAGAGGAGCAACCCCTACTATGTGTATGATGATGATGATGATGATGAT 886
 DB 792 CAAAACCTCAGAGAGAGGAGCAACCCCTACTATGTGTATGATGATGATGATGATGATGAT 851
 OY 887 CTGAGATATTAAGACATCCAGTCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 946
 DB 852 CTGAGATATTAAGACATCCAGTCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 911
 OY 947 GTGGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1006
 DB 912 GTGGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 971
 OY 1007 GTGGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1066
 DB 972 GTGGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1031
 OY 1067 GGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1126
 DB 1032 GGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1091
 OY 1127 TGCGGCTGCTCTTGTGCTTATGAAAGGTTTCTGACTCAGAAAGCTGTGGGAGTGC 1186
 DB 1092 TGCGGCTGCTCTTGTGCTTATGAAAGGTTTCTGACTCAGAAAGCTGTGGGAGTGC 1151
 OY 1187 ATGAATTTCTGAGGCTACCGCAGCAGTCAATGCAAGAAACAGAAACCTGTGCTG 1246
 DB 1152 ATGAATTTCTGAGGCTACCGCAGCAGTCAATGCAAGAAACAGAAACCTGTGCTG 1211
 OY 1247 GATCTTCTGCTTACCTGAGCTGCTGCTGAGCAATCACTAAGAGGTTGAGAGGCTTT 1306
 DB 1212 GATCTTCTGCTTACCTGAGCTGCTGCTGAGCAATCACTAAGAGGTTGAGAGGCTTT 1271
 OY 1307 GGTGACACCTGAGTGTGAGCCTGAGTCACTTCTGTAATGCTTGTAGGCTTGCTACAT 1366
 DB 1272 GGTGACACCTGAGTGTGAGCCTGAGTCACTTCTGTAATGCTTGTAGGCTTGCTACAT 1331
 OY 1367 CT 1368
 DB 1332 CT 1333

RESULT 5
 AAK91803
 ID AAK91803 standard; cDNA; 877 BP.
 XX
 AC AAK91803;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human cDNA 5'-end sequence, SEQ ID NO: 263.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP130094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-0114089.
 XX
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX

Db 508 TGATGCTACATGGAACCTGGTCCCGGAGACAGTCCAGCTCTTTTCAGTGGGAGAACAG 567
 QY 514 CCGGGCCCTGTGTCTGACCTGGCTGTGACATCATCCACAGATGAATCATCCCGG 573
 Db 568 CCGGGCCCTGTGTCTGACCTGGCTGTGACATCATCCACAGATGAATCATCCCGG 627
 QY 574 GGCCTACCTGACGGCTCCTGTGACAGTCAACAGCCCTTTGAGGTGTGTGAGAAAGCCCTGGC 633
 Db 628 GGCCTACCTGACGGCTCCTGTGACAGTCAACAGCCCTTTGAGGTGTGTGAGAAAGCCCTGGC 687
 QY 634 CTCTACAGCCAGCTGGTGTCAAGAGGTGGCTGTCTTCAACCTGCACTGCTCCCTCAACC 693
 Db 688 CTCTACAGCCAGCTGGTGTCAAGAGGTGGCTGTCTTCAACCTGCACTGCTCCCTCAACC 747
 QY 694 GCTGCTGAAGAACCCCTCCCTCAACCTGCAATATGTCAGCTGGTCCGGAGACCCCGGGC 753
 Db 748 GCTGCTGAAGAACCCCTCCCTCAACCTGCAATATGTCAGCTGGTCCGGAGACCCCGGGC 807
 QY 754 CGTGTCCGTTCCCGAGAACGACCAAGAGGAGATCTCATGATTTGACAGTGCATTTGAT 813
 Db 808 CGTGTCCGTTCCCGAGAACGACCAAGAGGAGATCTCATGATTTGACAGTGCATTTGAT 867
 QY 814 GGGGACAGTGAAGCAAAACTCAAGAAAGAGAGACCAACCTCACTATGATGAGAGGTAT 873
 Db 868 GGGGACAGTGAAGCAAAACTCAAGAAAGAGAGACCAACCTCACTATGATGAGAGGTAT 927
 QY 874 CTGCAAGAGCCAGCTGAGATCTCAAGAACCATCCAGTCTTGGCCAAAGGCCCTGCAAGA 933
 Db 928 CTGCAAGAGCCAGCTGAGATCTCAAGAACCATCCAGTCTTGGCCAAAGGCCCTGCAAGA 987
 QY 934 ACGTACCTGCTGTGGCTATGAGAGACCTGGCTGAGACCCCTGTGGCCAGACTTCCCG 993
 Db 988 ACGTACCTGCTGTGGCTATGAGAGACCTGGCTGAGACCCCTGTGGCCAGACTTCCCG 1047
 QY 994 AATGTATGAATTCGTGGGATTTGGAATCTTCCCATCTTCAAGCTGGGATGATATCAT 1053
 Db 1048 AATGTATGAATTCGTGGGATTTGGAATCTTCCCATCTTCAAGCTGGGATGATATCAT 1107
 QY 1054 CACCCGAGGCAAGGGGATGGGTGACACAGCTTTTCACACAAATGCCAGGATGCCCTTAA 1113
 Db 1108 CACCCGAGGCAAGGGGATGGGTGACACAGCTTTTCACACAAATGCCAGGATGCCCTTAA 1167
 QY 1114 TGTCTCCAGGCTTTGGGCTGTGTGTGCTTATGAAAAGTTTCTGACTTCAGAAAGC 1173
 Db 1168 TGTCTCCAGGCTTTGGGCTGTGTGTGCTTATGAAAAGTTTCTGACTTCAGAAAGC 1227
 QY 1174 CTGTGGCGATGACATGATTTGTGGGCTACCCGACAGTCTGATCTGAACAAGAACAGAG 1233
 Db 1228 CTGTGGCGATGACATGATTTGTGGGCTACCCGACAGTCTGATCTGAACAAGAACAGAG 1287
 QY 1234 AATCCTGTGTGGATTTCTGTCTACCTGAGCTGTCCCTGAGCAAAATCCACTAAGAGGG 1293
 Db 1288 AATCCTGTGTGGATTTCTGTCTACCTGAGCTGTCCCTGAGCAAAATCCACTAAGAGGG 1347
 QY 1294 TTGGAAGGCTTTGCTGCAACCTGGTGTGACCCCTCACTCACTTCTCTGATGCTTCTGA 1353
 Db 1348 TTGGAAGGCTTTGCTGCAACCTGGTGTGACCCCTCACTCACTTCTCTGATGCTTCTGA 1407
 QY 1354 GCCCTTGCTACATCTGAGCCTTAATACATGTCTGAGGTATCACACTGAGTGTGAGT 1413
 Db 1408 GCCCTTGCTACATCTGAGCCTTAATACATGTCTGAGGTATCACACTGAGTGTGAGT 1467
 QY 1414 TGTGTCCACAGTGTCTCAAGCAGAGACTTTTGTGTGCATGCTTGTGTAGAAACAG 1473
 Db 1468 TGTGTCCACAGTGTCTCAAGCAGAGACTTTTGTGTGCATGCTTGTGTAGAAACAG 1527
 QY 1474 ACTGGGAACTTATGTGAGCAGACATCCACAGTGAACAGAGGATTTGCTTCTTC 1533
 Db 1528 ACTGGGAACTTATGTGAGCAGACATCCACAGTGAACAGAGGATTTGCTTCTTC 1587
 QY 1534 TTTTCTGATCTTCTGCTGTGGGAGACTTCAGAGACTTTTGGCTGGAGGCTATTTAA 1593
 Db 1588 TTTTCTGATCTTCTGCTGTGGGAGACTTCAGAGACTTTTGGCTGGAGGCTATTTAA 1647

QY 1594 GCACGACACAGTATCAGTGAATTCATTAACCTCCCTGTCACATCTTGCCCAATG 1653
 Db 1648 GCACGACACAGTATCAGTGAATTCATTAACCTCCCTGTCACATCTTGCCCAATG 1707
 QY 1654 GGGAAATGATCTTTCACCAAGAGCTCACAGCATTTTTCACAGAGATGCCAATTTGAG 1713
 Db 1708 GGGAAATGATCTTTCACCAAGAGCTCACAGCATTTTTCACAGAGATGCCAATTTGAG 1767
 QY 1714 CCGTTGGATGCCCAATGGGATTCAGAGAGAGTGGGAACAAGTTGATGCTACTT 1773
 Db 1768 CCGTTGGATGCCCAATGGGATTCAGAGAGAGTGGGAACAAGTTGATGCTACTT 1826
 QY 1774 ATGAGCTTGACCAT-ACAGCTATGCGTATTCAGAAATTTGAACAAATCTCTG-ACAA 1831
 Db 1827 ATGAGCTTGACCATCAGAGTATGCGTATTCAGAAATTTGAACAAATCTCTGACAAA 1886
 QY 1832 AGAGCAAGCTTTAAGTTCAACAGAGTCCCTGGCTTGATTTGAATATCA-TTCCCTTTG 1890
 Db 1887 AGAGCAAGCTTTAAGTTCAACAGAGTCCCTGGCTTGATTTGAATATCACTTCCCTCTG 1946
 QY 1891 CATTTTCCCATTCATAGAAACCTTGAACCTGTGAACCTGCAATCTGTATTAATCAAA 1950
 Db 1947 CATTTTCCCATTCATAGAAACCTTGAACCTGTGAAC-TGCCATCTGTATTAATCAAA 2005
 QY 1951 TTCCCAATTAAG 1962
 Db 2006 TTCCCAATTAAG 2017

RESULT 4

AAS16947
 ID AAS16947 standard; cDNA; 1333 BP.

AC AAS16947;

DT 12-MAR-2002 (first entry)

XX Human L-selectin sulfotransferase-2 (LSST-2) cDNA.

KW Human: beta1,3gal: beta1,3-N-acetylgalactosaminyltransferase; MECA-79; ss;
 KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
 KW ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;
 KW allergic contact dermatitis; lymphoma; chronic pneumonia; LSST-2;
 KW delayed-type hypersensitivity reaction; hyperplastic thymus; antiulcer;
 KW antiinflammatory; antipsoriatic; antidiabetic; dermatological;
 KW antiallergic.

OS Homo sapiens.

XX Key Location/Qualifiers
 FH 111..1253
 FT /*tag= a
 FT /product= "Human LSST-2"

PN WO200185177-A1.

PD 15-NOV-2001.

PE 10-MAY-2001; 2001WO-US1452.

PR 11-MAY-2000; 2000US-0569320.

PA (BURN-) BURNHAM INST.

PI Fukuda M, Yeh J, Hirooka N;

DR WPI; 2002-075226/10.

DR P-PSDB; AAU11274.

PT New enzyme, useful for modifying acceptor molecule, comprises an
 PT isolated L-selectin sulfotransferase-2 that directs expression of
 PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or

```

Db      1099 TCACCCGAGGCAAGGGCATGGGTGACCACGCTTTCCACAAATGGCAGGATGCCCTTA 1158
QY      1113 ATGTCTCCAGGCGTTGGCGCTGGTCTTTGCGCTATGAAAGGTTTCTGACTTCAGAAAG 1172
Db      1159 ATGTCTCCAGGCGTTGGCGCTGGTCTTTGCGCTATGAAAGGTTTCTGACTTCAGAAAG 1218
QY      1173 CCTGTGGCGATGCGCATGAAATTTGCTGGGCTTACCGCCAGCTCAGATCTGAACAAGAAC 1232
Db      1219 CCTGTGGCGATGCGCATGAAATTTGCTGGGCTTACCGCCAGCTCAGATCTGAACAAGAAC 1278
QY      1233 GAAACCTGTGGTGGATCTTGTCTGCTACCTGGAGCTGCTCCCTGAGCAAAATCCACTAGAGG 1292
Db      1279 GAAACCTGTGGTGGATCTTGTCTGCTACCTGGAGCTGCTCCCTGAGCAAAATCCACTAGAGG 1338
QY      1293 GTTGAGAGAGGCTTTGGTGGCGACCTGGTGTGAGCCCTGAGCTCAGCTTCTCTGAATGCTTCTG 1352
Db      1339 GTTGAGAGAGGCTTTGGTGGCGACCTGGTGTGAGCCCTGAGCTCAGCTTCTCTGAATGCTTCTG 1398
QY      1353 AGCCTTGCTTACATCTCTGAGGCTTAACTACATGTCTGTGGGTATCACACTGAGTGTGAG 1412
Db      1399 AGCCTTGCTTACATCTCTGAGGCTTAACTACATGTCTGTGGGTATCACACTGAGTGTGAG 1458
QY      1413 TTGTGTCCACAGCTGCTCAAGCAGAGAGACTTTTGTGTCACATCTGTGTCTAGAAACA 1472
Db      1459 TTGTGTCCACAGCTGCTCAAGCAGAGAGACTTTTGTGTCACATCTGTGTCTAGAAACA 1518
QY      1473 GACTGGGGAACCTTATGTAGAGACACATCCACAGTAACAGGATATGCTCTTCTT 1532
Db      1519 GACTGGGGAACCTTATGTAGAGACACATCCACAGTAACAGGATATGCTCTTCTT 1578
QY      1533 CTTTCTTATCTCTCTGCTGGGCACTTCAAGAACCTTTGTGGCTTGAGGCTTATTA 1592
Db      1579 CTTTCTTATCTCTCTGCTGGGCACTTCAAGAACCTTTGTGGCTTGAGGCTTATTA 1638
QY      1593 AGCAGCACACAGTATCAGTGAATGATTCATTAACCTCCCTGCTCCACATCTTCCCAAT 1652
Db      1639 AGCAGCACACAGTATCAGTGAATGATTCATTAACCTCCCTGCTCCACATCTTCCCAAT 1698
QY      1653 GGGGAATGATCTTTCACCAAGAGCTCACAGCATTTTTCACAGATGCGAATCTGA 1712
Db      1699 GGGGAATGATCTTTCACCAAGAGCTCACAGCATTTTTCACAGATGCGAATCTGA 1758
QY      1713 GCCCTTGAGTCCCAATGGGATTCAGAGAGAGTGGGAACAGGTTGATGCTTACT 1772
Db      1759 GCCCTTGAGTCCCAATGGGATTCAGAGAGAGTGGGAACAGGTTGATGCTTACT 1817
QY      1773 TATGACCTTGACCAT-ACAGCTATCGTATCAGAAATATGAAACAAATCTCTG-ACAA 1830
Db      1818 TATGACCTTGACCATCACAAGCTATCGTATCAGAAATATGAAACAAATCTCTGACAA 1877
QY      1831 AAGAGCAGCTCTTAAGTTCAAGAGTGGCTGGGCTTGAATTTGAATATCA-TTCCCTT 1889
Db      1878 AAGAGCAGCTCTTAAGTTCAAGAGTGGCTGGGCTTGAATTTGAATATCACTCCCTCT 1937
QY      1890 GCATTTCCATTAAGTAAAGAACTTGAAGCTGTAAGTATGCAATCTGTAATATCTTAA 1949
Db      1938 GCATTTCCATTAAGTAAAGAACTTGAAGCTGTAAGTATGCAATCTGTAATATCTTAA 1996
QY      1950 ATTCCCAATTAAGTCTGTTTGAATGCTCCCTTTTATGCTCTTAATATTAAGAGTA 2009
Db      1997 ATTCCCAATTAAGTCTGTTTGAATGCTCCCTTTTATGCTCTTAATATTAAGAGTA 2056
QY      2010 AATGTTCAAT 2018
Db      2057 AATGTTCAAT 2065

```

RESULT 3
AAZ20792
ID AAZ20792 standard; DNA; 2032 BP.

XX AC AAZ20792;
XX

```

DT      08-DEC-1999 (first entry)
DE      Human glycosyl sulfotransferase-3 coding sequence.
XX
XX      Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
KW      selectin binding interaction; Inflammation; lymphocyte homing; human;
KW      secondary lymph organ; ss.
XX
XX      Homo sapiens.
OS
XX      MO9949018-A1.
XX
XX      30-SEP-1999.
XX
XX      26-FEB-1999; 99WO-US04316.
XX
XX      20-MAR-1998; 98US-0045284.
XX
XX      12-NOV-1998; 98US-0190911.
XX
XX      (REGC ) UNIV CALIFORNIA.
XX      (SYNT ) SYNTAX USA INC.
XX
XX      Bisturup A, Rosen SD, Tangemann K, Hemmerlich S;
XX      WPI, 1999-580442/49.
XX      P-P-SDB; AAY39918.
XX
XX      Human and murine glycosyl sulfotransferase 3 and related
XX      polynucleotides
XX
XX      Claim 4; Fig 1; 59pp; English.
XX
XX      This sequence encodes the human glycosyl sulfotransferase-3 (GST-3) of
XX      the invention. The nucleic acid sequences, probes and primers derived
XX      from these, proteins and antibodies are useful in detecting homologues.
XX      The sequences, antibodies and methods are useful in the diagnosis and
XX      treatment of diseases associated with selectin binding interactions,
XX      including conditions associated with or resulting from the homing of
XX      leukocytes to sites of inflammation and the normal homing of lymphocytes
XX      to secondary lymph organs.
XX
XX      Sequence 2032 BP; 468 A; 569 C; 490 G; 505 T; 0 other;
SQ

```

Query Match 87.4%; Score 1784.8; DB 20; Length 2032;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1850; Conservative 0; Mismatches 17; Indels 5; Gaps 5;

```

QY      94 AAGCCCAAGCAGCAAGGCTTCCACTTCAGACACATGCTACTGCTTAAAAAATGAAGCT 153
Db      148 AAGCCCGCTCTTGCAGAGTCTTCCACTTCAGACACATGCTACTGCTTAAAAAATGAAGCT 207
QY      154 CCTGCTGTTTCTGTTTCCAGATGGCAATCTTGCTTATTTCTTCACATGTACAGCA 213
Db      208 CCGTGTGTTTCTGTTTCCAGATGGCAATCTTGCTTATTTCTTCACATGTACAGCA 267
QY      214 CAACATCAGCTCCCTCTATGAGGACACAGCCGAGCGAGCAGCTGCTGTCTG 273
Db      268 CAACATCAGCTCCCTCTATGAGGACACAGCCGAGCGAGCAGCTGCTGTCTG 327
QY      274 TTCTGCGGCTGCGCTCTTCTTGTGGGCGAGCTTTTGGCAGCAGCCAGATTTT 333
Db      328 TTCTGCGGCTGCGCTCTTCTTGTGGGCGAGCTTTTGGCAGCAGCCAGATTTT 387
QY      334 CTACATGATGAGCCCGCTGCGACGTGATGATGATCTTCAAGCAGACAGCCCGCTGAT 393
Db      388 CTACATGATGAGCCCGCTGCGACGTGATGATGATCTTCAAGCAGACAGCCCGCTGAT 447
QY      394 GCTGCAATGAGCTGTGCGGAGATGTATACGGCGCTCTTCTGTGCGACATGAGCTCTT 453
Db      448 GCTGCAATGAGCTGTGCGGAGATGTATACGGCGCTCTTCTGTGCGACATGAGCTCTT 507
QY      454 TGATGCTACATGAGAACCTGTGCTCCCGGAGACAGTCCAGCTCTTCAATGGAGAACG 513

```

|||||
Db 1834 CATTTCCTCCATACATAGACATGCTGACCTGTGAAGC-TGCCATCTGTATATCTATAAA 1892
Qy 1951 TTCCCAATAGAGTCTCTTGTAGAAATGTCCTTTTATGCTTCTTATATATAGCAGTAA 2010
Db 1893 TTCCCAATAGAGTCTCTTGTAGAAATGTCCTTTTATGCTTCTTATATATAGCAGTAA 1952
Qy 2011 ATGTTCATTTTATGAGATCCTAA 2034
Db 1953 ATGTTCATTTTATGAGATCCTAA 1976
RESULT 2
AA294211
ID AA294211 standard; cDNA; 2065 BP.
AA294211;
XX 19-JUN-2000 (first entry)
Df Human transferase TRNSFS-11 cDNA clone 2617407CB1.
XX
Df
XX
XX
KW Transferase; TRNSFS-11; human; antitumor; cell proliferation;
inflammation; gastrointestinal disorder; developmental disorder;
genetic disorder; neurological disorder; reproductive disorder;
smooth muscle disorder; immunological disorder; gene therapy;
diagnosis; N-acetylglucosamine 6-O-sulfotransferase; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 174..1334
FT /tag- a
FN MO200014251-A2.
XX
XX 16-MAR-2000.
XX
PF 09-SEP-1999; 99WO-US20989.
XX
XX 10-SEP-1998; 98US-0150657.
PR 04-NOV-1998; 98US-0186779.
PR 11-MAY-1999; 99US-0133642.
XX
XX (INCY-) INCYTE PHARM INC.
PI Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;
PI Hillman JL, Azimzal Y;
XX WPI: 2000-256996/22.
DR P-PSDB: AAY79219.
XX
XX Human transferase proteins useful for preventing, diagnosing and
treating cancers and developmental, gastrointestinal, genetic,
immunological, neurological, reproductive and smooth muscle disorders.
XX
XX Claim 9; Page 104-105; 113pp; English.
XX
XX The present sequence is that of cDNA clone 2617407CB1 encoding
human transferase TRNSFS-11 (see AAY79219), 1 of 15 claimed human
transferase proteins of the invention (see AAY79209-23). The clone
was isolated from gall bladder cDNA library GBLANOT01. TRNSFS-11
is expressed in dermatologic and gastrointestinal tissues,
especially those associated with inflammation and cell
proliferation. It shows homology to mouse N-acetylglucosamine
6-O-sulfotransferase. A polynucleotide comprising nucleotides
264-333 or 1272-1331 of the present sequence can be used as a DNA
probe. The new human transferases and polynucleotides can be used
in the diagnosis, prevention and treatment (including gene therapy
and antitense therapy) of cancer, developmental disorders,
gastrointestinal disorders, genetic disorders, immunological
disorders, neurological disorders, reproductive disorders, and
smooth muscle disorders.

XX SQ Sequence 2065 BP; 466 A; 575 C; 491 G; 533 T; 0 other;
Query Match 88.7%; Score 1813; DB 21; Length 2065;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1902; Conservative 0; Mismatches 20; Indels 7; Gaps 7;
Qy 94 AAGCCCAAGCACAAGATGCTTCCACTTACAGCAATAGTACTGCTTAAAAAATGAAGCT 153
Db 140 AAGCCCGCTTGCAAGGCTTCCACTTACAGCAATAGTACTGCTTAAAAAATGAAGCT 199
Qy 154 CCGTCTGTTTCTGTTTCCAGATGCGCATCTTGCGTATTTCTTCCATGATACAGCA 213
Db 200 CCGTCTGTTTCTGTTTCCAGATGCGCATCTTGCGTATTTCTTCCATGATACAGCA 259
Qy 214 CAACATCAGCTCCCGTATGAAAGGACAGCCGAGGAGGACGAGCGTGGTCTGTC 273
Db 260 CAACATCAGCTCCCGTATGAAAGGACAGCCGAGGAGGACGAGCGTGGTCTGTC 319
Qy 274 TTCTGAGCGCTCTGCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 333
Db 320 TTCTGAGCGCTCTGCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 379
Qy 334 CTACCTGATGAGAGCCCGCTGACGCTGTGATGACCTTTCAGACAGACACCGCTGAT 393
Db 380 CTACCTGATGAGAGCCCGCTGACGCTGTGATGACCTTTCAGACAGACACCGCTGAT 439
Qy 394 GCTGCACATGAGCTGAGCGGAGATCATACGGGCGCTCTTCTTCTTCTTCTTCTTCTTCTT 453
Db 440 GCTGCACATGAGCTGAGCGGAGATCATACGGGCGCTCTTCTTCTTCTTCTTCTTCTTCTT 499
Qy 454 TGATGCCATACATGAACTGTGTCCCGGAGACAGTCCAGCTCTTCTTCTTCTTCTTCTTCTT 513
Db 500 TGATGCCATACATGAACTGTGTCCCGGAGACAGTCCAGCTCTTCTTCTTCTTCTTCTTCTT 559
Qy 514 CCGGCGCTGTGTTCTGACCTGCTGTGATCATCCACAGATGAATCATCCCGG 573
Db 560 CCGGCGCTGTGTTCTGACCTGCTGTGATCATCCACAGATGAATCATCCCGG 618
Qy 574 GGGTCACAGGAGGCTGTCAGATCAACAGCCCTTGAAGTGG-TGAGAAAGGCTGCG 632
Db 619 GGGTCACAGGAGGCTGTCAGATCAACAGCCCTTGAAGTGG-TGAGAAAGGCTGCG 678
Qy 633 GCTCTACAGCAGCTGTGCTCAAGAGTGCCTTCTTCACTGACGTCCCTTACC 692
Db 679 GCTCTACAGCAGCTGTGCTCAAGAGTGCCTTCTTCACTGACGTCCCTTACC 738
Qy 693 CCGTCTGAAAGACCCCTCCCTCAACTGCTATGCTGACCTGCTCCGAGACCCCGGG 752
Db 739 CCGTCTGAAAGACCCCTCCCTCAACTGCTATGCTGACCTGCTCCGAGACCCCGGG 798
Qy 753 CCGTCTGCTTCCCGAGACGACCAAGAGATCTCATATTGACAGTGGCATTTGGA 812
Db 799 CCGTCTGCTTCCCGAGACGACCAAGAGATCTCATATTGACAGTGGCATTTGGA 858
Qy 813 TGGGGCAGCATGAGCAAAACTCAAGAGAGGACCAACCTTACTATGTGATGAGGTCA 872
Db 859 TGGGGCAGCATGAGCAAAACTCAAGAGAGGACCAACCTTACTATGTGATGAGGTCA 918
Qy 873 TCTGCAAGACGACCTGAGATCTTCAAGACATTCAGTCTTGGCCCAAGGCTTGGCA 932
Db 919 TCTGCAAGACGACCTGAGATCTTCAAGACATTCAGTCTTGGCCCAAGGCTTGGCA 978
Qy 933 AAGCTTACCTGCTGTGAGATGAGGACCTGCTGAGGCGCTGAGGCGGACGCTTCC 992
Db 979 AAGCTTACCTGCTGTGAGATGAGGACCTGCTGAGGCGCTGAGGCGGACGCTTCC 1038
Qy 993 GAATGATGAAATTCGTGGGATTTGAAATCTTCCCATCTTACAGACCTGAGGTGATACA 1052
Db 1039 GAATGATGAAATTCGTGGGATTTGAAATCTTCCCATCTTACAGACCTGAGGTGATACA 1098
Qy 1053 TCACCCGAGGAGGAGGAGTGGGTGACACAGCTTTCACACAAATGCGAGGATCCCTTA 1112
Db 1112 TCACCCGAGGAGGAGGAGTGGGTGACACAGCTTTCACACAAATGCGAGGATCCCTTA

XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8; SEQ ID NO 2816; 1380bp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesized by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a full length
CC human cDNA of the invention.
CC Note: The sequence data in this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 1979 BP; 451 A; 542 C; 474 G; 512 T; 0 other;

Query Match 90.9%; Score 1856.8; DB 22; Length 1979;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1922; Conservative 0; Mismatches 17; Indels 5; Gaps 5;

QY 94 AAGCCAGAGCCACAGGCTTCCACTTCAAGACACATGCTACTGCTTAAAAAATGAAGCT 153
DB 35 AAGCCAGGCTGTGCAAGAGTCTTCCACTTCAAGACACATGCTACTGCTTAAAAAATGAAGCT 94
QY 154 CCGTGTGTTGTTGTTCCAGATGAGCCATCTGCTGCTATCTTCCACATGTAAGCA 213
DB 95 CCGTGTGTTGTTGTTCCAGATGAGCCATCTGCTGCTATCTTCCACATGTAAGCA 154
QY 214 CAACATCAGCTCCCTCTATGAGGACAGCCGAGCGCATCAGTGTGTTGTC 273
DB 155 CAACATCAGCTCCCTCTATGAGGACAGCCGAGCGCATCAGTGTGTTGTC 214
QY 274 TTCTGTGCGCTGTGCTCTTCTTGTGGGACAGCTTTTGGGACAGCCAGATGTTT 333
DB 215 TTCTGTGCGCTGTGCTCTTCTTGTGGGACAGCTTTTGGGACAGCCAGATGTTT 274
QY 334 CTACATGATGAGACCGGCGTGGACAGTGGATGATGATCTTCAAGCAGAGCCGCTGAT 393
DB 275 CTACATGATGAGACCGGCGTGGACAGTGGATGATGATCTTCAAGCAGAGCCGCTGAT 334
QY 394 GCTGACATGCTGTGCGGAGATCTGATACGGGCGCTTCTTGTGACATGAGCGTCT 453
DB 335 GCTGACATGCTGTGCGGAGATCTGATACGGGCGCTTCTTGTGACATGAGCGTCT 394
QY 454 TGATGCTATGATGAGACCTGTGCTCCCGGAGACAGTCCAGCTTTCAAGTGGAGAACAG 513
DB 395 TGATGCTATGATGAGACCTGTGCTCCCGGAGACAGTCCAGCTTTCAAGTGGAGAACAG 454
QY 514 CCGGCGCTGTGTTGTGACACTGCTGTGACATCATCCCAAGATGAATATCTCCCGG 573
DB 455 CCGGCGCTGTGTTGTGACACTGCTGTGACATCATCCCAAGATGAATATCTCCCGG 514
QY 574 GGTCTACTGAGGCTCTGTGACATCAAGCCCTTGAAGTGTGAGAGAGGCGTCCG 633
DB 515 GGTCTACTGAGGCTCTGTGACATCAAGCCCTTGAAGTGTGAGAGAGGCGTCCG 574
QY 634 CTCTTACAGCCAGTGTGCTCAAGAGAGTGGCTTCTTCAACTGACGTCCCTTAACC 693
DB 575 CTCTTACAGCCAGTGTGCTCAAGAGAGTGGCTTCTTCAACTGACGTCCCTTAACC 634
QY 694 GGTCTGAGAGAGAGCTCTCCCTCAACCTGATCTGACAGCTGTGCGGAGAGCCGCGG 753
DB 635 GGTCTGAGAGAGAGCTCTCCCTCAACCTGATCTGACAGCTGTGCGGAGAGCCGCGG 694
QY 754 CGTGTTCCTTCCGAGAGAGCAAAAGGAGATCTCATGATTGACAGTGCATTTGAT 813
DB 695 CGTGTTCCTTCCGAGAGAGCAAAAGGAGATCTCATGATTGACAGTGCATTTGAT 754

QY 814 GGGGACATGAGCAAAAACTCAAGAGAGAGACCAACCTACTATGATGAGGCTCAT 873
DB 755 GGGGACATGAGCAAAAACTCAAGAGAGAGACCAACCTACTATGATGAGGCTCAT 814
QY 874 CTGCCAAAGCCAGCTGGAGATCTCAAGACATCTCCTTCCAGAGCCCTGACAGA 933
DB 815 CTGCCAAAGCCAGCTGGAGATCTCAAGACATCTCCTTCCAGAGCCCTGACAGA 874
QY 934 AGCTTCTGCTGTGCGGATGAGAGACCGTGCAGGCCCTGTGGCCAGACTTCCG 993
DB 875 AGCTTCTGCTGTGCGGATGAGAGACCGTGCAGGCCCTGTGGCCAGACTTCCG 934
QY 994 AATGATGATTTGTTGGATTTGGAATCTTCCCATCTTCAAGACCTGGGTGATTAAT 1053
DB 935 AATGATGATTTGTTGGATTTGGAATCTTCCCATCTTCAAGACCTGGGTGATTAAT 994
QY 1054 CACCCAGGCAAGGAGATGGGTACCAAGCTTTCCACAAATGCCAGGATGCCCTTA 1113
DB 995 CACCCAGGCAAGGAGATGGGTACCAAGCTTTCCACAAATGCCAGGATGCCCTTA 1054
QY 1114 TGTCTCCAGAGCTTGGCGCTGTGCTTGTGCGCTTGAAGGTTCTGACTGAGAAAG 1173
DB 1055 TGTCTCCAGAGCTTGGCGCTGTGCTTGTGCGCTTGAAGGTTCTGACTGAGAAAG 1114
QY 1174 CTGTGCGATGCCATGAATTTGCTGGGCTACCGCCACGTGATCTGAACAGACAGAG 1233
DB 1115 CTGTGCGATGCCATGAATTTGCTGGGCTACCGCCACGTGATCTGAACAGACAGAG 1174
QY 1234 AATCGTGTGCGATCTTCTGTACCTGAGATGTCCTGAGCAATCCATCAAGAGAG 1293
DB 1175 AATCGTGTGCGATCTTCTGTACCTGAGATGTCCTGAGCAATCCATCAAGAGAG 1234
QY 1294 TTGAGAGAGCTTGTGCTGACACCTGCTGTGACAGCTTCTCTGATGCTTGA 1353
DB 1235 TTGAGAGAGCTTGTGCTGACACCTGCTGTGACAGCTTCTCTGATGCTTGA 1294
QY 1354 GCTTGTCTATCTCTGAGCTTAACTATCATCTGTGGGTATCATCACTGAGTGAAT 1413
DB 1295 GCTTGTCTATCTCTGAGCTTAACTATCATCTGTGGGTATCATCACTGAGTGAAT 1354
QY 1414 TGTGTCCACACGTGCTCAAGCAGAGAGACTTGTGTCTCATGCTGTGTAGAAAACAG 1473
DB 1355 TGTGTCCACACGTGCTCAAGCAGAGAGACTTGTGTCTCATGCTGTGTAGAAAACAG 1414
QY 1474 ACTGGGAACTTATGTGAGACACATCCACCAAGTGAAGAGGATTTGCTTCTTC 1533
DB 1415 ACTGGGAACTTATGTGAGACACATCCACCAAGTGAAGAGGATTTGCTTCTTC 1474
QY 1534 TTTTCTTGAATCTTCTGTGCGGACAGACTTGTGCGCTGGAGGCTTATTA 1593
DB 1475 TTTTCTTGAATCTTCTGTGCGGACAGACTTGTGCGCTGGAGGCTTATTA 1534
QY 1594 GCACGACACAGTATGAGTGAATTTGATCCATAAACCCTGTCACATCTTCCCAATG 1653
DB 1535 GCACGACACAGTATGAGTGAATTTGATCCATAAACCCTGTCACATCTTCCCAATG 1594
QY 1654 GGGATGATCTTTCACAAAGAGCTCACAGCATTTTCCACAGAGATGGAATTTGAG 1713
DB 1595 GGGATGATCTTTCACAAAGAGCTCACAGCATTTTCCACAGAGATGGAATTTGAG 1654
QY 1714 CCTTGGAGTTTCCCATGAGATTCAGAGAGAGTGGGAAACAAGTTGATGCTTACT 1773
DB 1655 CCTTGGAGTTTCCCAT -GGATTCAAGGAGAGAGTGGGAAACAAGTTGATGCTTACT 1713
QY 1774 ATGAGCTTGACAT -ACAGTATCGGTAATCAGAAATATGAACAAATCTCTG -ACAA 1831
DB 1714 ATGAGCTTGACATCAGATCTATCGTATCAGAAATATGAACAAATCTCTGACAA 1773
QY 1832 AGAGCAAGCTCTTAAGTCAAGAGTGGCTTGAATTTGAATTA -TTTCCCTTG 1890
DB 1774 AGAGCAAGCTCTTAAGTCAAGAGTGGCTTGAATTTGAATTA -TTTCCCTTG 1833
QY 1891 CATTTTCCATTTACATAGAAAACCTTGGACTGTGAACCTTGGCATCTGTTAATACTAAA 1950

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 10:39:13 ; Search time 507.533 Seconds
(without alignments)
10866.196 Million cell updates/sec

Title: US-09-645-078-1
Perfect score: 2043
Sequence: 1 gaattcattgttgcgtgta.....tgggatacctataaaaaaaa 2043

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_19Jun03:.*
1: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1856.8	90.9	1979	22	AAK94229 Human full-length
2	1813	88.7	2065	21	AAZ94211 Human transferase
3	1784.8	87.4	2032	20	AAZ20792 Human glycosyl sul
4	1262	61.8	1333	24	AAK16947 Human l-selectin s
5	765.6	37.4	877	22	AAK91803 Human cDNA 5'-end
6	763.6	37.4	877	22	AAK93921 Human cDNA clone r
7	743.2	36.4	1926	20	AAZ20793 Mouse glycosyl sul
8	547.2	26.8	2988	21	AAK76156 Human ORFX ORF1711

9	480.2	23.5	517	24	ABK54724 Human colon cancer
10	478.8	23.4	505	24	ABV89280 Human colon cancer
11	412.8	20.2	1647	24	AAZ24670 Human drug metabol
12	412.8	20.2	1694	22	AAK02700 Human glycosyl sul
13	412.8	20.2	2544	24	ABN89506 Human corneal N-ac
14	412.8	20.2	160552	22	AAK02697 Human glycosyl sul
15	398.4	19.5	2044	22	AAK02699 Human glycosyl sul
16	398.4	19.5	2170	22	AAK02698 Human glycosyl sul
17	387.2	19.0	1937	24	AAK16948 Murine intestinal-
18	387.2	19.0	1989	22	AAK02696 Mouse glycosyl sul
19	376.2	18.4	548	22	AAK92588 Human cDNA 3'-end
20	336.2	16.5	48436	24	ABN89533 Human corneal N-ac
21	294.4	14.4	389	24	ABK54794 Human colon cancer
22	144.8	7.1	7099	25	ACC46083 Human dthp enzyme
23	141.6	6.9	2156	19	AAV21200 Glycosaminoglycan
24	141	6.9	668	24	ABK021506 Oligonucleotide fo
25	141	6.9	668	24	ABK021507 Oligonucleotide fo
26	135.2	6.6	1458	19	AAV36418 Keratan sulphate 6
27	135.2	6.6	2190	24	AAK41280 Human chondroitin
28	135.2	6.6	2415	25	ABX08787 Angiogenesis-assoc
29	128.4	6.3	2354	18	AAK45037 Chick chondroitin
30	123.8	6.1	3029	22	AAH17922 Human cDNA sequenc
31	119.4	5.8	2393	24	AAK61699 Lung small cell ca
32	119.4	5.8	2409	20	AAK87821 Human N-acetylgluc
33	109.4	5.4	524	24	ABK021510 Oligonucleotide fo
34	109.4	5.4	524	20	ABK021511 Oligonucleotide fo
35	108	5.3	2150	20	AAK87820 Mouse N-acetylgluc
36	106.4	5.2	524	24	ABK021512 Oligonucleotide fo
37	106.4	5.2	524	24	ABK021513 Oligonucleotide fo
38	104.4	5.1	668	24	ABK021508 Oligonucleotide fo
39	104.4	5.1	668	24	ABK021509 Oligonucleotide fo
40	91	4.5	576	24	ABK045408 Oligonucleotide fo
41	91	4.5	576	24	ABK045409 Oligonucleotide fo
42	85.2	4.2	2165	25	ABK21794 Human chondroitin
43	68	3.3	576	24	ABK045406 Oligonucleotide fo
44	68	3.3	576	24	ABK045407 Oligonucleotide fo
45	62.6	3.1	731	20	AAK24563 Human lung tumor a

ALIGNMENTS

RESULT 1	
AAK94229	
ID	AAK94229 standard; cDNA; 1979 BP.
XX	
AC	AAK94229;
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human full-length cDNA, SEQ ID NO: 2816.
XX	
KW	Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX	
OS	Homo sapiens.
XX	
PN	EP1130094-A2.
XX	
PD	05-SEP-2001.
XX	
PF	07-JUL-2000; 2000EP-0114089.
XX	
PR	08-JUL-1999; 98JP-0194486.
XX	
PR	11-JAN-2000; 2000JP-0118774.
XX	
PR	02-MAY-2000; 2000JP-0183765.
XX	
PA	(HELI-) HELIX RES INST.
XX	
PI	Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
XX	
PI	Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX	
DR	WPI, 2001-524255/58.
DR	P-PSDB; AAK93309.